

Rigorous Dynamics of Expectation-Propagation-Based Signal Recovery from Unitarily Invariant Measurements

Keigo Takeuchi, *Member, IEEE*

Abstract—Signal recovery from unitarily invariant measurements is investigated in this paper. A message-passing algorithm is formulated on the basis of expectation propagation (EP). A rigorous analysis is presented for the dynamics of the algorithm in the large system limit, where both input and output dimensions tend to infinity while the compression rate is kept constant. The main result is the justification of state evolution (SE) equations conjectured by Ma and Ping. This result implies that the EP-based algorithm achieves the Bayes-optimal performance derived by Takeda et al. in 2006 via a non-rigorous tool in statistical physics, when the compression rate is larger than a threshold. The proof is based on an extension of a conventional conditioning technique for the standard Gaussian matrix to the case of the Haar matrix.

Index Terms—Compressed sensing, expectation propagation, unitarily invariant measurements, state evolution, Haar matrices.

I. INTRODUCTION

A. Motivation

CONSIDER the recovery problem of an N -dimensional signal vector \mathbf{x} from a compressed noisy measurement vector $\mathbf{y} \in \mathbb{C}^M$ ($M \leq N$) [2], [3],

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}. \quad (1)$$

In (1), $\mathbf{A} \in \mathbb{C}^{M \times N}$ denotes a known measurement matrix. The signal vector \mathbf{x} is an unknown sparse¹ vector that is composed of independent and identically distributed (i.i.d.) elements. The noise vector $\mathbf{w} \in \mathbb{C}^M$ is independent of the other random variables. The goal of compressed sensing is to recover the sparse vector \mathbf{x} from the knowledge about \mathbf{y} and \mathbf{A} , as well as the statistics of all random variables.

A breakthrough for signal recovery is to construct message-passing (MP) that is Bayes-optimal in the large system limit, where the input and output dimensions N and M tend to infinity while the compression rate $\delta = M/N$ is kept constant.

Manuscript received April *, 2017. The author was in part supported by the Grant-in-Aid for Exploratory Research (JSPS KAKENHI Grant Number 15K13987), Japan. The material in this paper will be presented in part at 2017 IEEE International Symposium on Information Theory, Aachen, Germany, Jun. 2017 [1].

K. Takeuchi is with the Department of Electrical and Electronic Information Engineering, Toyohashi University of Technology, Aichi 441-8580, Japan (e-mail: takeuchi@ee.tut.ac.jp).

¹ In this paper, a signal $\mathbf{x} \in \mathbb{R}$ is called sparse if the Rényi information dimension [4] of \mathbf{x} is smaller than 1. If \mathbf{x} is zero with probability $1 - p$, the information dimension is at most p . If \mathbf{x} is discrete, it is zero.

The origin of this approach dates back to the Thouless-Anderson-Palmer (TAP) equation [5] in statistical physics. Motivated by the TAP approach, Kabashima [6] proposed an MP algorithm based on approximate belief propagation (BP) in the context of code-division multiple-access (CDMA) systems with i.i.d. measurement matrices. When the compression rate is larger than the so-called BP threshold [7], the BP-based algorithm was numerically shown to achieve the Bayes-optimal performance in the large system limit, which was originally conjectured by Tanaka [8] via the replica method—a non-rigorous tool in statistical physics, and proved in [9], [10] for i.i.d. Gaussian measurements. However, Kabashima [6] presented no rigorous analysis on the convergence property of the BP-based algorithm.

In order to resolve lack of a rigorous proof, approximate message-passing (AMP) was proposed in [11] and proved in [12] to achieve the optimal performance for i.i.d. Gaussian measurements, when the compression rate is larger than the BP threshold. Spatially coupled measurement matrices are required for achieving the optimal performance in the whole regime [7], [13]–[15]. However, it is recognized that AMP fails to converge when the i.i.d. assumption of measurement matrices is broken [16], unless damping [17] is employed.

As solutions to this convergence issue, since Oppor and Winther's pioneering work [18, Appendix D], as well as [19], several algorithms have been proposed on the basis of expectation propagation (EP) [20], expectation consistent (EC) approximations [18], [21], [22], S-transform [23], approximate BP [24], or turbo principle [25]–[28]. The EP-based algorithm [20] is systematically derived from Minka's EP framework [29] by approximating the posterior distribution of \mathbf{x} with factorized Gaussian distributions. The EC-based algorithms [18], [21], [22] are iterative algorithms for solving a fixed point (FP) of the EC free energy. An algorithm in [23] is derived via the S-transform of $\mathbf{A}^H \mathbf{A}$. Rangan *et al.* [24] considered an EP-like approximation of the BP algorithm on a factor graph with vector-valued nodes. The algorithms [25]–[28] based on turbo principle are derived from a few heuristic assumptions. Interestingly, the algorithms in [18], [20], [22], [24], [27] are essentially equivalent, with the exception of [21], [23]. In this paper, these algorithms for signal recovery are simply referred to as EP-based algorithms, since we follow the EP-based derivation in [20].

Ma *et al.* [26], [27] derived state evolution (SE) equations

of an EP-based algorithm under two heuristic assumptions. By investigating the properties of the SE equations, they conjectured that, for unitarily invariant measurement matrices, the FPs of the SE equations are the same as those of an asymptotic energy function that describes the Bayes-optimal performance in the large system limit, which were derived in [30] via the replica method. In other words, the EP-based algorithm was conjectured to achieve the optimal performance in the large system limit, when the compression rate is larger than the BP threshold. Since the algorithm attempts to solve a FP of the EC free energy [18], it is conjectured that the FPs of the EC free energy correspond to those of the Bayes-optimal one for unitarily invariant measurement matrices. The purpose of this paper is to present a rigorous proof for the conjecture.

B. Proof Strategy

The proof strategy is based on a conditioning technique used in [12]. A challenging part in the proof is to evaluate the distribution of an estimation error in each iteration conditioned on estimation errors in all preceding iterations. Bayati and Montanari [12] evaluated the conditional distribution via the distribution of the measurement matrix \mathbf{A} conditioned on the estimation errors in all preceding iterations. When linear detection is employed as part of MP, the conditional distribution of \mathbf{A} can be regarded as the posterior distribution of \mathbf{A} given linear, noiseless, and compressed observations of \mathbf{A} , determined by the estimation errors in all preceding iterations. For i.i.d. Gaussian measurement matrices, it is well known that the posterior distribution is also Gaussian. The proof in [12] heavily relies on this well-known fact.

In order to present our proof strategy, assume $M = N$, and that \mathbf{A} is a Haar matrix [31], [32], which is uniformly distributed on the space of all possible $N \times N$ unitary matrices. Under appropriate coordinate rotations in the row and column spaces of \mathbf{A} , it is possible to show that the linear, noiseless, and compressed observation of \mathbf{A} is equivalent to observing *part* of the elements in \mathbf{A} . Since any Haar matrix is bi-unitarily invariant [31], the distribution of \mathbf{A} after the coordinate rotations is the same as the original one. Thus, evaluating the conditional distribution of \mathbf{A} reduces to analyzing the conditional distribution of a Haar matrix given part of its elements. This argument was implicitly used in [12].

Evaluation of this conditional distribution is a technically challenging part in this paper, while this part is not required for i.i.d. Gaussian measurements. Intuitively, conditioning on a small part of the elements can be ignored, since the number of conditioned elements is $\mathcal{O}(N)$ —much smaller than the number of all elements N^2 in the large system limit. In order to prove this intuition, we use coordinate rotations to reveal the statistical structure of the conditional Haar matrix. We know that a Haar matrix has similar properties to an i.i.d. Gaussian matrix as $N \rightarrow \infty$. In particular, a finite number of linear combinations of the elements in a Haar matrix were proved to converge toward jointly Gaussian-distributed random variables in the large system limit [33], [34]. Note that the classical central limit theorem cannot be used, since the elements of

a Haar matrix are not independent. Using this asymptotic similarity between Haar and i.i.d. Gaussian matrices, we prove the intuition.

C. Related Work

A similar paper [24] was posted on the arXiv a few months before posting the first version [35] of this paper, of which a short paper will be presented in [1]. Interestingly, the two papers share the common proof strategy based on [12]. However, there are a few differences between them.

First of all, we consider complex-valued systems, while the posted paper [24] addressed real-valued systems. The main difference is that a probabilistic approach is used in this paper, while a deterministic approach based on pseudo-Lipschitz continuity was considered in [24]. The deterministic approach only provides results averaged over all elements of \mathbf{x} . On the other hand, the probabilistic one allows us to obtain decoupling results for individual elements of the signal vector \mathbf{x} —stronger than the averaged results.

D. Contributions

The main contribution is the rigorous justification of the SE equations for the EP-based algorithm, conjectured in [27]. More precisely, we derive SE equations for individual elements of the signal vector in the large system limit. This implies the achievability of the Bayes-optimal performance conjectured in [30], when the compression rate is larger than the BP threshold, while the converse theorem is still open, i.e. there are no algorithms outperforming the EP-based algorithm in the large system limit.

The technical novelty is in an extension of the conditioning technique in [12] for i.i.d. Gaussian measurement matrices to the case of Haar matrices. This paper presents a constructive proof for the conditional distribution of a Haar matrix, which the correctness of the same result was proved in [24]. The proof strategy of the main theorem is applicable to any MP algorithm for signal recovery from unitarily invariant measurements, such as the AMP, unless the algorithm contains nonlinear processing in the measurement vector \mathbf{y} , e.g. quantization [28].

E. Organization

The remainder of this paper is organized as follows: After summarizing the notation used in this paper, Section II presents the definition of unitarily invariant matrices and a few technical results associated with Haar matrices. In Section III, we introduce assumptions used throughout this paper, and then formulate an EP-based algorithm. The main result is presented in Section IV, and proved in Section V. Several technical results are proved in appendices.

F. Notation

The proper complex Gaussian distribution with mean \mathbf{m} and covariance $\mathbf{\Sigma}$ is denoted by $\mathcal{CN}(\mathbf{m}, \mathbf{\Sigma})$. For random variables X and Y , the notation $X \stackrel{\text{a.s.}}{=} Y$ means that X is almost surely equal to Y . Similarly, $\stackrel{\text{a.s.}}{\rightarrow}$, $\stackrel{\text{a.s.}}{\geq}$, and $\stackrel{\text{a.s.}}{\leq}$ indicates that \rightarrow , \geq ,

and \leq hold almost surely. The notation $X \sim Y$ means that X follows the same distribution as Y , while $X \xrightarrow{d} Y$ is used to represent that X converges in distribution to Y . The notation $X|_Y$ indicates that we focus on the conditional distribution of X given Y .

The notation $\mathbf{o}(1)$ denotes a vector of which the Euclidean norm converges almost surely toward zero in the large system limit. For a vector $\mathbf{v} \in \mathbb{C}^N$, we write the n th element of \mathbf{v} as v_n , in which the index n runs from 0 to $N - 1$. For $k \in \mathbb{N}$, the family $\mathfrak{N}_k = \{N \subset \{0, \dots, N - 1\} : |N| = k\}$ of sets is composed of all possible subsets of the indices with cardinality k . For a subset $N \subset \{0, \dots, N - 1\}$, the vector \mathbf{x}_N consists of the elements $\{x_n : n \in N\}$. For a scalar function $f : \mathbb{C} \rightarrow \mathbb{C}$, we introduce a convention in which $f(\mathbf{v})$ denotes the vector obtained by the component-wise application of f to \mathbf{v} , i.e. $[f(\mathbf{v})]_n = f(v_n)$.

For a complex number $z \in \mathbb{C}$ and a matrix $M \in \mathbb{C}^{M \times N}$, the complex conjugate, transpose, and the conjugate transpose are denoted by z^* , M^T , and M^H . We write the (m, n) th element of M as M_{mn} . When M is Hermitian, $\lambda_{\min}(M)$ represents the minimum eigenvalue of M . For $M \geq N$, $\mathcal{U}_{M \times N}$ denotes the space of all possible $M \times N$ matrices with orthonormal columns, while $\mathcal{U}_{M \times N}$ for $M < N$ represents the space of all possible $M \times N$ matrices with orthonormal rows. When $M = N$ holds, $\mathcal{U}_{M \times N}$ is written as \mathcal{U}_M , which is the space of all possible $M \times M$ unitary matrices.

We write the singular-value decomposition (SVD) of M as

$$M = \Phi_M (\Sigma_M, O) \Psi_M^H \quad (2)$$

for $M \leq N$, with $\Phi_M \in \mathcal{U}_M$ and $\Psi_M \in \mathcal{U}_N$. Furthermore, Σ_M is an $M \times M$ positive semi-definite diagonal matrix. The unitary matrix Ψ_M is partitioned as $\Psi_M = (\Psi_M^{\parallel}, \Psi_M^{\perp})$, in which $\Psi_M^{\parallel} \in \mathcal{U}_{N \times M}$ is composed of the first M columns of Ψ_M , while $\Psi_M^{\perp} \in \mathcal{U}_{N \times (N-M)}$ consists of the remaining columns. For $M > N$, we write the SVD of M as

$$M = \Phi_M \begin{pmatrix} \Sigma_M \\ O \end{pmatrix} \Psi_M^H, \quad (3)$$

with $\Phi_M \in \mathcal{U}_M$ and $\Psi_M \in \mathcal{U}_N$. Furthermore, Σ_M is an $N \times N$ positive semi-definite diagonal matrix. The unitary matrix $\Phi_M = (\Phi_M^{\parallel}, \Phi_M^{\perp})$ is partitioned in the same manner as for $M \leq N$.

When M is full rank, the pseudo-inverse of M is denoted by $M^\dagger = (M^H M)^{-1} M^H \in \mathbb{C}^{N \times M}$ for $M > N$. Let P_M^{\parallel} denote the orthogonal projection matrix onto the space spanned by the columns of M . We have $P_M^{\parallel} = \Phi_M^{\parallel} (\Phi_M^{\parallel})^H = M M^\dagger$. The projection matrix P_M^{\perp} onto the orthogonal complement is given by $P_M^{\perp} = I_M - P_M^{\parallel}$. For $M \leq N$, we define $M^\dagger = M^H (M M^H)^{-1}$, $P_M^{\parallel} = \Psi_M^{\parallel} (\Psi_M^{\parallel})^H = M^\dagger M$, and $P_M^{\perp} = I_N - P_M^{\parallel}$.

II. PRELIMINARY

The purpose of this section is to present two technical results: the strong law of large numbers and the central limit theorem for a Haar matrix. The two results correspond to [12, Lemma 2 (a) and (b)] for an i.i.d. Gaussian matrix. We start with the definition of a Haar matrix.

Definition 1: A unitary random matrix $U \in \mathcal{U}_n$ is called a Haar matrix if U is uniformly distributed on \mathcal{U}_n .

An important property of a Haar matrix is bi-unitary invariance [32]—used throughout this paper.

Definition 2: A random matrix M is called bi-unitarily invariant if $M \sim U M V$ holds for all deterministic unitary matrices U and V .

We first present the strong law of large numbers associated with a Haar matrix. The elements of a Haar matrix are not independent of each other, so that we utilize the strong law of large numbers for dependent random variables.

Theorem 1 (Lyons [36]): Let $\{X_i\}_{i=1}^\infty$ denote a sequence of complex random variables with finite second moments, and define $S_n = \sum_{i=1}^n X_i$. The strong law of large numbers for $T_n = (S_n - \mathbb{E}[S_n])/n$ holds, i.e. $\lim_{n \rightarrow \infty} T_n \stackrel{\text{a.s.}}{=} 0$, if the following assumption holds:

$$\sum_{n=1}^{\infty} \frac{\sqrt{\mathbb{V}[S_n]}}{n^2} < \infty. \quad (4)$$

Proof: See [36, Theorem 6]. ■

The following lemma is the strong law of large numbers associated with a Haar matrix.

Lemma 1: Suppose that $V \in \mathcal{U}_N$ is a Haar matrix. Let $\mathbf{a} \in \mathbb{C}^N$ and $\mathbf{b} \in \mathbb{C}^N$ denote random vectors that are independent of V and satisfy $\lim_{N \rightarrow \infty} N^{-1} \|\mathbf{a}\|^2 \stackrel{\text{a.s.}}{=} 1$, $\lim_{N \rightarrow \infty} N^{-1} \|\mathbf{b}\|^2 \stackrel{\text{a.s.}}{=} 1$, and $\lim_{N \rightarrow \infty} N^{-1} \mathbf{b}^H \mathbf{a} \stackrel{\text{a.s.}}{=} C$. Furthermore, we define a Hermitian matrix $D \in \mathbb{C}^{N \times N}$ such that D is independent of V , and that $N^{-1} \text{Tr}(D^2)$ is almost surely convergent as $N \rightarrow \infty$. Then,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbf{b}^H V \mathbf{a} \stackrel{\text{a.s.}}{=} 0, \quad (5)$$

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbf{b}^H V^H D V \mathbf{a} \stackrel{\text{a.s.}}{=} C \lim_{N \rightarrow \infty} \frac{1}{N} \text{Tr}(D). \quad (6)$$

Proof: See Appendix A. ■

We next present the central limit theorem associated with a Haar matrix.

Theorem 2 (Chatterjee and Meckes [34]): Let $V \in \mathcal{U}_N$ denote a Haar matrix. For any $k \in \mathbb{N}$, suppose that k matrices $\{W_i \in \mathbb{C}^{N \times N}\}$ are independent of V , and satisfy $\text{Tr}(W_i W_j^H) = N \delta_{i,j}$ for all $i, j = 1, \dots, k$, in which $\delta_{i,j}$ denotes the Kronecker delta. Then, the vector $\mathbf{a} = (\text{Tr}(V W_1), \dots, \text{Tr}(V W_k))^T$ converges in distribution to the standard complex Gaussian vector $\mathbf{z} \sim \mathcal{CN}(0, I_k)$ as $N \rightarrow \infty$.

Proof: See [34, Theorem 4.6]. ■

Lemma 1 and Theorem 2, as well as Theorem 1, will be used in the proof of the main theorem.

III. SYSTEM MODEL

A. Assumptions

Assumptions on the measurement model (1) are presented.

Assumption 1: The signal vector \mathbf{x} is composed of zero-mean i.i.d. non-Gaussian elements with unit variance and finite fourth moments.

The i.i.d. assumption for \mathbf{x} is implicitly used in the derivation of an EP-based algorithm. We require no additional

assumptions for the prior distribution of each element to prove the main theorem, whereas it is practically important to postulate some prior distribution indicating the sparsity of \mathbf{x} .

Definition 3: A Hermitian random matrix \mathbf{M} is called unitarily invariant if $\mathbf{M} \sim \mathbf{U}\mathbf{M}\mathbf{U}^H$ holds for any deterministic unitary matrix \mathbf{U} .

Assumption 2: The measurement matrix \mathbf{A} has the following properties:

- $\mathbf{A}^H \mathbf{A}$ is unitarily invariant.
- The empirical eigenvalue distribution of $\mathbf{A}\mathbf{A}^H$ converges almost surely to a deterministic distribution $\rho(\lambda)$ with a finite fourth moment in the large system limit.

We write the SVD of \mathbf{A} as

$$\mathbf{A} = \mathbf{U}(\boldsymbol{\Sigma}, \mathbf{O})\mathbf{V}^H, \quad (7)$$

with $\mathbf{U} \in \mathcal{U}_M$ and $\mathbf{V} \in \mathcal{U}_N$. Furthermore, $\boldsymbol{\Sigma}$ is an $M \times M$ positive semi-definite diagonal matrix. From Assumption 2, \mathbf{V} is a Haar matrix and independent of $\mathbf{U}\boldsymbol{\Sigma}$ [32].

Assumption 3: Let $\mathbf{D} \in \mathbb{C}^{M \times M}$ denote any Hermitian matrix such that \mathbf{D} is independent of \mathbf{w} , and that $N^{-1}\text{Tr}(\mathbf{D}^2)$ is almost surely convergent as $N \rightarrow \infty$. Then, the noise vector \mathbf{w} satisfies

$$\lim_{M \rightarrow \infty} \frac{1}{M} \mathbf{w}^H \mathbf{U} \mathbf{D} \mathbf{U}^H \mathbf{w} \stackrel{\text{a.s.}}{=} \lim_{M \rightarrow \infty} \frac{\sigma^2}{M} \text{Tr}(\mathbf{D}). \quad (8)$$

Assumption 3 implies that σ^2 corresponds to the noise power $\sigma^2 \stackrel{\text{a.s.}}{=} \lim_{M \rightarrow \infty} M^{-1} \|\mathbf{w}\|^2$ per element, by selecting $\mathbf{D} = \mathbf{I}_M$. Assumption 3 is satisfied if \mathbf{w} is unitarily invariant, e.g. $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$, or if \mathbf{U} is a Haar matrix and independent of \mathbf{D} .

B. Expectation Propagation

We start with an MP algorithm proposed in [27]. Let the detector postulate that the noise vector \mathbf{w} in (1) is a circularly symmetric complex Gaussian (CSCG) random vector with covariance $\sigma^2 \mathbf{I}_M$. This postulation needs not be consistent with the true distribution of \mathbf{w} .

As derived in Appendix B, the MP algorithm for this case is based on EP and composed of two modules. In iteration t , a first module—called module A—calculates the *extrinsic* mean $\mathbf{x}_{A \rightarrow B}^t$ and variance $v_{A \rightarrow B}^t$ of the signal vector \mathbf{x} from $\mathbf{x}_{B \rightarrow A}^t$ and $v_{B \rightarrow A}^t$ provided by the other module—called module B.

$$\mathbf{x}_{A \rightarrow B}^t = \mathbf{x}_{B \rightarrow A}^t + \gamma_t \mathbf{W}^t (\mathbf{y} - \mathbf{A} \mathbf{x}_{B \rightarrow A}^t), \quad (9)$$

$$v_{A \rightarrow B}^t = \gamma_t - v_{B \rightarrow A}^t. \quad (10)$$

In the initial iteration $t = 0$, the prior mean $\mathbf{x}_{B \rightarrow A}^0 = \mathbf{0}$ and variance $v_{B \rightarrow A}^0 = N^{-1} \mathbb{E}[\|\mathbf{x}\|^2] = 1$ are used.

In (9), the linear minimum mean-square error (LMMSE) filter $\mathbf{W}^t \in \mathbb{C}^{N \times M}$ is given by

$$\mathbf{W}^t = \mathbf{A}^H \left(\sigma^2 \mathbf{I}_M + v_{B \rightarrow A}^t \mathbf{A} \mathbf{A}^H \right)^{-1}. \quad (11)$$

The normalization coefficient γ_t in (9) is defined as

$$\frac{1}{\gamma_t} = \lim_{M \rightarrow \infty} \frac{1}{N} \text{Tr}(\mathbf{W}^t \mathbf{A}) \stackrel{\text{a.s.}}{=} \frac{1}{\gamma(v_{B \rightarrow A}^t)} \quad (12)$$

due to Assumption 2, with

$$\frac{1}{\gamma(v)} = \int \frac{\delta \lambda}{\sigma^2 + v \lambda} d\rho(\lambda), \quad (13)$$

where $\rho(\lambda)$ denotes the asymptotic eigenvalue distribution of $\mathbf{A}\mathbf{A}^H$ in the large system limit. The coefficient γ_t keeps the orthogonality between estimation errors in the two modules.

On the other hand, module B computes the minimum mean-square error (MMSE) estimator and the MMSE of x_n

$$\tilde{\eta}_t(x_{n,A \rightarrow B}^t) = \mathbb{E}[x_n | x_{n,A \rightarrow B}^t], \quad (14)$$

$$\text{MMSE}(v_{A \rightarrow B}^t) = \mathbb{E}[|x_n - \tilde{\eta}_t(x_{n,A \rightarrow B}^t)|^2], \quad (15)$$

given the virtual additive white Gaussian noise (AWGN) observation,

$$x_{n,A \rightarrow B}^t = x_n + z_n^t, \quad z_n^t \sim \mathcal{CN}(0, v_{A \rightarrow B}^t). \quad (16)$$

From Assumption 1, note that the MMSE is independent of n . If a termination condition is satisfied, module B outputs $\tilde{\eta}_t(\mathbf{x}_{A \rightarrow B}^t)$ as an estimate of \mathbf{x} . Otherwise, module B feeds the extrinsic mean $\mathbf{x}_{B \rightarrow A}^{t+1}$ and variance $v_{B \rightarrow A}^{t+1}$ of \mathbf{x} back to module A, given by

$$\mathbf{x}_{B \rightarrow A}^{t+1} = \eta_t(\mathbf{x}_{A \rightarrow B}^t), \quad (17)$$

$$\frac{1}{v_{B \rightarrow A}^{t+1}} = \frac{1}{\text{MMSE}(v_{A \rightarrow B}^t)} - \frac{1}{v_{A \rightarrow B}^t}, \quad (18)$$

where the decision function $\eta_t : \mathbb{C} \rightarrow \mathbb{C}$ is defined as

$$\eta_t(x) = v_{B \rightarrow A}^{t+1} \left(\frac{\tilde{\eta}_t(x)}{\text{MMSE}(v_{A \rightarrow B}^t)} - \frac{x}{v_{A \rightarrow B}^t} \right). \quad (19)$$

Remark 1: The decision function $\eta_t(x)$ is zero for all $x \in \mathbb{C}$ if and only if $x_n \sim \mathcal{CN}(0, 1)$ holds. We have postulated Assumption 1 to let the decision function be non-constant.

The following lemma presents an important property of the decision function η_t in module B.

Lemma 2 (Ma and Ping [27]): Suppose that $z_n^t \sim \mathcal{CN}(0, v_{A \rightarrow B}^t)$ is an independent CSCG random variable with variance $v_{A \rightarrow B}^t$. Then,

$$\lim_{\epsilon \rightarrow 0} \mathbb{E}_{z_n^t} [(z_n^t)^* \eta_t(x_n + \epsilon + z_n^t)] = 0, \quad (20)$$

$$\lim_{\epsilon \rightarrow 0} \mathbb{E}_z [(z_n^t)^* \tilde{\eta}_t(x_n + \epsilon + z_n^t)] = \text{MMSE}(v_{A \rightarrow B}^t). \quad (21)$$

Proof: See Appendix C for the proof based on [27]. ■

Lemma 2 will be used to prove the orthogonality between estimation errors in the two modules.

Remark 2: As considered in [24], [27], we can replace the decision function η_t with another suboptimal function that satisfies the properties in Lemma 2. Such a replacement may be important when the true prior distribution of the signal elements is unknown. However, the replacement provides no influences on the proof of the main theorem, with the exception of the stability of the derived SE equations. Thus, we only consider the optimal decision function η_t , without loss of generality.

C. Error Recursion

An error recursion for the EP-based algorithm is formulated to analyze the convergence property. Let $\mathbf{h}_t = \mathbf{x} - \mathbf{x}_{A \rightarrow B}^t$ and $\mathbf{q}_t = \mathbf{x} - \mathbf{x}_{B \rightarrow A}^t$ denote the estimation errors for the extrinsic estimates in modules A and B, respectively. Substituting the system model (1) into (9), and using the SVD (7) and (17), we obtain the error recursion

$$\mathbf{b}_t = \mathbf{V}^H \mathbf{q}_t, \quad (22)$$

$$\mathbf{m}_t = \mathbf{b}_t - \gamma_t \tilde{\mathbf{W}}_t \{(\boldsymbol{\Sigma}, \mathbf{O}) \mathbf{b}_t + \tilde{\mathbf{w}}\}, \quad (23)$$

$$\mathbf{h}_t = \mathbf{V} \mathbf{m}_t, \quad (24)$$

$$\mathbf{q}_{t+1} = \mathbf{q}_0 - \eta_t (\mathbf{q}_0 - \mathbf{h}_t), \quad (25)$$

with $\tilde{\mathbf{w}} = \mathbf{U}^H \mathbf{w}$. In (23), the linear filter $\tilde{\mathbf{W}}_t$ is given by

$$\tilde{\mathbf{W}}_t = (\boldsymbol{\Sigma}, \mathbf{O})^H (\sigma^2 \mathbf{I}_M + v_{B \rightarrow A}^t \boldsymbol{\Sigma}^2)^{-1}. \quad (26)$$

Furthermore, we define $\eta_{-1}(\cdot) = 0$ to obtain $\mathbf{q}_0 = \mathbf{x}$.

In analyzing the convergence property, we focus on the distribution of the estimation error \mathbf{h}_t conditioned on the preceding iteration history. Thus, it is useful to represent the error recursion in the matrix form. Define

$$\begin{aligned} \mathbf{Q}_t &= (\mathbf{q}_0, \dots, \mathbf{q}_{t-1}) \in \mathbb{C}^{N \times t}, \\ \mathbf{B}_t &= (\mathbf{b}_0, \dots, \mathbf{b}_{t-1}) \in \mathbb{C}^{N \times t}, \\ \mathbf{M}_t &= (\mathbf{m}_0, \dots, \mathbf{m}_{t-1}) \in \mathbb{C}^{N \times t}, \\ \mathbf{H}_t &= (\mathbf{h}_0, \dots, \mathbf{h}_{t-1}) \in \mathbb{C}^{N \times t}. \end{aligned} \quad (27)$$

The error recursion is represented as

$$\mathbf{V}^H \mathbf{Q}_t = \mathbf{B}_t, \quad (28)$$

$$\mathbf{M}_t = \mathbf{G}_t(\mathbf{B}_t), \quad (29)$$

$$\mathbf{V} \mathbf{M}_t = \mathbf{H}_t, \quad (30)$$

$$\mathbf{Q}_{t+1} = \mathbf{F}_t(\mathbf{H}_t, \mathbf{q}_0), \quad (31)$$

where the τ th columns of $\mathbf{G}_t(\mathbf{B}_t)$ and $\mathbf{F}_t(\mathbf{H}_t, \mathbf{q}_0)$ are equal to the right-hand sides (RHSs) of (23) and (25) for $t = \tau$, respectively.

IV. MAIN RESULT

We define the individual mean-square error (MSE) for the extrinsic estimate in module B as

$$\text{mse}_{n,B \rightarrow A}^t \stackrel{\text{a.s.}}{=} \lim_{M=\delta N \rightarrow \infty} \mathbb{E}[|q_{t,n}|^2] \quad (32)$$

in the large system limit. Furthermore, let mse_n^t denote the individual MSE for the estimate (14) in the EP-based algorithm in the large system limit, given by

$$\text{mse}_n^t = \lim_{M=\delta N \rightarrow \infty} \mathbb{E}[|x_n - \tilde{\eta}_t(x_{n,A \rightarrow B}^t)|^2]. \quad (33)$$

The following theorem is the main result of this paper, which describes the rigorous dynamics of the EP-based algorithm in the large system limit.

Theorem 3 (State Evolution): Define deterministic SE equations as

$$\overline{\text{mse}}_{A \rightarrow B}^t = \gamma(\overline{\text{mse}}_{B \rightarrow A}^t) - \overline{\text{mse}}_{B \rightarrow A}^t, \quad (34)$$

TABLE I
NOTATIONAL CONVENTIONS FOR $t = 0$.

$\mathcal{X}_{0,0} = \{\mathbf{Q}_1\}, \mathcal{X}_{0,1} = \{\mathbf{Q}_1, \mathbf{B}_1, \mathbf{M}_1 \mathbf{M}_1 = \mathbf{G}_1(\mathbf{B}_1)\},$ $\mathbf{Q}_0 = \mathbf{O}, \mathbf{B}_0 = \mathbf{O}, \mathbf{M}_0 = \mathbf{O}, \mathbf{H}_0 = \mathbf{O}, \mathbf{M}_0^\dagger = \mathbf{O}, \mathbf{Q}_0^\dagger = \mathbf{O},$ $\mathbf{P}_{\mathbf{M}_0}^\perp = \mathbf{I}_N, \mathbf{P}_{\mathbf{Q}_0}^\perp = \mathbf{I}_N, \boldsymbol{\alpha}_0 = \mathbf{0}, \boldsymbol{\beta}_0 = \mathbf{0},$ $\boldsymbol{\Phi}_{\mathbf{V}_{10}^{0,t'}}^\perp = \mathbf{I}_{N-t'}, \boldsymbol{\Phi}_{\mathbf{M}_0}^\perp = \mathbf{I}_N, \mathcal{V}_{0,1} = \{\mathbf{V}_{01}^{0,1}\}, \bar{\mathbf{V}}_{11}^{0,1} = \mathbf{O}.$

$$\frac{1}{\overline{\text{mse}}_{B \rightarrow A}^{t+1}} = \frac{1}{\text{MMSE}(\overline{\text{mse}}_{A \rightarrow B}^t)} - \frac{1}{\overline{\text{mse}}_{A \rightarrow B}^t}, \quad (35)$$

with $\overline{\text{mse}}_{B \rightarrow A}^0 = 1$, in which $\gamma(\cdot)$ and $\text{MMSE}(\cdot)$ are given in (13) and (15), respectively. Then, the individual MSEs $\text{mse}_{n,B \rightarrow A}^t$ and mse_n^t are equal to

$$\text{mse}_{n,B \rightarrow A}^t = \overline{\text{mse}}_{B \rightarrow A}^t, \quad (36)$$

$$\text{mse}_n^t = \text{MMSE}(\overline{\text{mse}}_{A \rightarrow B}^t) \quad (37)$$

in the large system limit for all n .

Remark 3: In order to simplify the proof of Theorem 3, the individual MSE for the extrinsic estimate in module A is not analyzed in this paper. However, our proof strategy can be applied to justifying that the individual MSE for module A converges to $\overline{\text{mse}}_{A \rightarrow B}^t$ in the large system limit.

Corollary 1: Define the MSEs averaged over the elements of \mathbf{x} as

$$\text{amse}_{B \rightarrow A}^t = \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbb{E}[\|\mathbf{q}_t\|^2], \quad (38)$$

$$\text{amse}^t = \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbb{E}[\|\mathbf{x} - \tilde{\eta}_t(\mathbf{x}_{A \rightarrow B}^t)\|^2]. \quad (39)$$

Then, the MSEs $\text{amse}_{B \rightarrow A}^t$ and amse^t are equal to $\overline{\text{mse}}_{B \rightarrow A}^t$ and $\text{MMSE}(\overline{\text{mse}}_{A \rightarrow B}^t)$ in the large system limit, respectively, which are defined in Theorem 3.

Corollary 1 was originally conjectured in [27], and implies that the EP-based algorithm predicts the exact dynamics of the extrinsic variances in the large system limit. Compare (10) and (18) to the SE equations. The FPs of the SE equations were proved in [27] to correspond to those of an asymptotic energy function that describes the Bayes-optimal performance—derived in [30] via the replica method. Thus, the Bayes-optimal performance derived in [30] is achievable when the SE equations have a unique FP, or equivalently when the compression rate δ is larger than the BP threshold.

We shall introduce several notations to present a general theorem, of which a corollary is Theorem 3. The random variables in the error recursions (28)–(31) are divided into three groups: \mathbf{V} , $\Theta = \{\boldsymbol{\Sigma}, \tilde{\mathbf{w}}\}$, and

$$\begin{aligned} \mathcal{X}_{t,t'} &= \left\{ \mathbf{Q}_{t+1}, \mathbf{B}_{t'}, \mathbf{M}_{t'}, \mathbf{H}_t \mid \mathbf{B}_{t'}^H \mathbf{M}_t = \mathbf{Q}_{t'}^H \mathbf{H}_t, \right. \\ &\quad \left. \mathbf{M}_{t'} = \mathbf{G}_{t'}(\mathbf{B}_{t'}), \mathbf{Q}_{t+1} = \mathbf{F}_t(\mathbf{H}_t, \mathbf{q}_0) \right\}, \end{aligned} \quad (40)$$

for $t' = t$ or $t' = t + 1$, while we define $\mathcal{X}_{0,0} = \{\mathbf{Q}_1\}$ and $\mathcal{X}_{0,1} = \{\mathbf{Q}_1, \mathbf{B}_1, \mathbf{M}_1 | \mathbf{M}_1 = \mathbf{G}_1(\mathbf{B}_1)\}$. See Table I for the notational conventions used in this paper.

The set Θ is fixed throughout this paper. Thus, conditioning on Θ is omitted. The set $\mathcal{X}_{t,t}$ describes the history of all preceding iterations just before updating (22), while $\mathcal{X}_{t,t+1}$

represents the history just before updating (24). Note that the condition $\mathbf{B}_{t'}^H \mathbf{M}_t = \mathbf{Q}_{t'}^H \mathbf{H}_t$ is a constraint imposing $\mathbf{V} \in \mathcal{U}_N$, and follows from (28) and (30). In order to investigate the dynamics of the error recursions, the distribution of the Haar matrix \mathbf{V} conditioned on $\mathcal{X}_{t,t'}$ is analyzed.

Let $\mathbf{m}_t^\perp = \mathbf{P}_{\mathbf{M}_t}^\perp \mathbf{m}_t$. Since $\mathbf{m}_t^\parallel = \mathbf{m}_t - \mathbf{m}_t^\perp$ is in the space spanned by the columns of \mathbf{M}_t , we have $(\mathbf{m}_t^\parallel)^H \mathbf{m}_t^\perp = 0$. Furthermore, \mathbf{m}_t^\parallel is represented as $\mathbf{m}_t^\parallel = \mathbf{M}_t \boldsymbol{\alpha}_t$, with $\boldsymbol{\alpha}_t = \mathbf{M}_t^\dagger \mathbf{m}_t \in \mathbb{C}^t$. Similarly, we define $\mathbf{q}_{t'}^\perp = \mathbf{P}_{\mathbf{Q}_{t'}}^\perp \mathbf{q}_{t'}$ and $\mathbf{q}_{t'}^\parallel = \mathbf{q}_{t'} - \mathbf{q}_{t'}^\perp = \mathbf{Q}_{t'} \boldsymbol{\beta}_{t'}$, with $\boldsymbol{\beta}_{t'} = \mathbf{Q}_{t'}^\dagger \mathbf{q}_{t'} \in \mathbb{C}^{t'}$.

For notational convenience, we define $\mathbf{Q}_0 = \mathbf{O}$, $\mathbf{B}_0 = \mathbf{O}$, $\mathbf{M}_0 = \mathbf{O}$, $\mathbf{H}_0 = \mathbf{O}$, $\mathbf{M}_0^\dagger = \mathbf{O}$, $\mathbf{Q}_0^\dagger = \mathbf{O}$, $\boldsymbol{\alpha}_0 = \mathbf{0}$, and $\boldsymbol{\beta}_0 = \mathbf{0}$. These definitions imply that $\mathbf{P}_{\mathbf{M}_0}^\perp = \mathbf{I}_N$ and $\mathbf{P}_{\mathbf{Q}_0}^\perp = \mathbf{I}_N$ hold.

Theorem 4: The following properties hold for each iteration $\tau = 0, 1, \dots$:

- (a) Each element $q_{\tau+1,n}$ in $\mathbf{q}_{\tau+1}$ has finite fourth moments. More precisely, for all n

$$\lim_{M=\delta N \rightarrow \infty} \mathbb{E} [|q_{\tau+1,n}|^4] < \infty. \quad (41)$$

For all $\tau' \leq \tau + 1$, the following limit exists:

$$\zeta_{\tau+1,\tau'} \stackrel{\text{a.s.}}{=} \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{q}_{\tau'}^H \mathbf{q}_{\tau+1}. \quad (42)$$

For some constant $C > 0$,

$$\liminf_{M=\delta N \rightarrow \infty} \lambda_{\min} \left(\frac{1}{N} \mathbf{M}_{\tau+1}^H \mathbf{M}_{\tau+1} \right) \stackrel{\text{a.s.}}{>} C, \quad (43)$$

$$\liminf_{M=\delta N \rightarrow \infty} \lambda_{\min} \left(\frac{1}{N} \mathbf{Q}_{\tau+2}^H \mathbf{Q}_{\tau+2} \right) \stackrel{\text{a.s.}}{>} C. \quad (44)$$

- (b) Let $\{\mathbf{z}_\tau \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)\}$ denote a sequence of independent standard complex Gaussian vectors that are independent of \mathbf{V} . Define

$$\tilde{\mathbf{b}}_\tau = \mathbf{B}_\tau \boldsymbol{\beta}_\tau + \mathbf{M}_\tau \mathbf{o}(1) + \mathbf{B}_\tau \mathbf{o}(1) + \mu_\tau^{1/2} \mathbf{z}_\tau, \quad (45)$$

$$\tilde{\mathbf{h}}_\tau = \mathbf{H}_\tau \boldsymbol{\alpha}_\tau + \mathbf{Q}_{\tau+1} \mathbf{o}(1) + \mathbf{H}_\tau \mathbf{o}(1) + \nu_\tau^{1/2} \mathbf{z}_\tau \quad (46)$$

with

$$\mu_\tau \stackrel{\text{a.s.}}{=} \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \|\mathbf{q}_\tau^\perp\|^2, \quad (47)$$

$$\nu_\tau \stackrel{\text{a.s.}}{=} \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \|\mathbf{m}_\tau^\perp\|^2. \quad (48)$$

Then, for any $k \in \mathbb{N}$

$$\mathbf{b}_{\tau,\mathcal{N}} |_{\Theta, \mathcal{X}_{\tau,\tau}} \xrightarrow{d} \tilde{\mathbf{b}}_{\tau,\mathcal{N}}, \quad (49)$$

$$\mathbf{h}_{\tau,\mathcal{N}} |_{\Theta, \mathcal{X}_{\tau,\tau+1}} \xrightarrow{d} \tilde{\mathbf{h}}_{\tau,\mathcal{N}} \quad (50)$$

hold for all subsets $\mathcal{N} \in \mathfrak{N}_k$ in the large system limit.

- (c) Let $\boldsymbol{\omega} \in \mathbb{C}^N$ denote any vector that is independent of \mathbf{V} , and satisfies $\lim_{N \rightarrow \infty} N^{-1} \|\boldsymbol{\omega}\|^2 \stackrel{\text{a.s.}}{=} 1$. Suppose that \mathbf{D} is any $N \times N$ Hermitian matrix such that \mathbf{D} depends only on $\boldsymbol{\Sigma}$, and that $N^{-1} \text{Tr}(\mathbf{D}^2)$ is almost surely convergent as $N \rightarrow \infty$. Then, for all $\tau' \leq \tau$ and $\tau'' \leq \tau + 1$

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{b}_{\tau'}^H \boldsymbol{\omega} \stackrel{\text{a.s.}}{=} 0, \quad (51)$$

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{b}_{\tau'}^H \mathbf{D} \mathbf{b}_{\tau} \stackrel{\text{a.s.}}{=} \zeta_{\tau,\tau'} \lim_{N \rightarrow \infty} \frac{1}{N} \text{Tr}(\mathbf{D}), \quad (52)$$

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{b}_{\tau'}^H \mathbf{m}_\tau \stackrel{\text{a.s.}}{=} 0, \quad (53)$$

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{m}_{\tau'}^H \mathbf{m}_\tau \stackrel{\text{a.s.}}{=} \gamma_{\tau,\tau'} - \zeta_{\tau,\tau'}, \quad (54)$$

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{h}_{\tau'}^H \mathbf{h}_\tau \stackrel{\text{a.s.}}{=} \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{m}_{\tau'}^H \mathbf{m}_\tau, \quad (55)$$

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{h}_\tau^H \mathbf{q}_{\tau''} \stackrel{\text{a.s.}}{=} 0, \quad (56)$$

with

$$\gamma_{t,t'} = \gamma_t \gamma_{t'} \int \frac{\delta \lambda (\sigma^2 + \zeta_{t,t'} \lambda)}{(\sigma^2 + v_{\mathbf{B} \rightarrow \mathbf{A}}^t \lambda)(\sigma^2 + v_{\mathbf{B} \rightarrow \mathbf{A}}^{t'} \lambda)} d\rho(\lambda). \quad (57)$$

- (d) The individual MSEs (33) and (32) for the posterior and extrinsic estimates (14) and (17) in module B coincide with the MMSE (15) and extrinsic variance (18) in the large system limit,

$$\text{mse}_n^\tau \stackrel{\text{a.s.}}{=} \text{MMSE}(v_{\mathbf{A} \rightarrow \mathbf{B}}^\tau), \quad (58)$$

$$\text{mse}_{n,\mathbf{B} \rightarrow \mathbf{A}}^{\tau+1} \stackrel{\text{a.s.}}{=} v_{\mathbf{B} \rightarrow \mathbf{A}}^{\tau+1}. \quad (59)$$

Proof: See Section V. ■

Ma and Ping [27, Assumption 1] postulated that (50) holds for the set of all indices $\mathcal{N} = \{0, \dots, N-1\}$. The assumption is too strong to be justified. In fact, the proof of Theorem 4 implies that the assumption may not be correct, since it is impossible to let $k = N$ in Theorem 2. However, the weaker property (b) is sufficient to prove Theorem 3.

Proof of Theorem 3: From (10), (18), (34), and (35), as well as from $\overline{\text{mse}}_{\mathbf{B} \rightarrow \mathbf{A}}^0 = v_{\mathbf{B} \rightarrow \mathbf{A}}^0 = 1$, we note that $\overline{\text{mse}}_{\mathbf{A} \rightarrow \mathbf{B}}^t = v_{\mathbf{A} \rightarrow \mathbf{B}}^t$ and $\overline{\text{mse}}_{\mathbf{B} \rightarrow \mathbf{A}}^t = v_{\mathbf{B} \rightarrow \mathbf{A}}^t$ hold for all t . Thus, Theorem 3 follows from (58) and (59). ■

V. PROOF OF THEOREM 4

A. Technical Lemmas

We need to evaluate the two distributions $p(\mathbf{m}_t, \mathbf{b}_t | \Theta, \mathcal{X}_{t,t})$ and $p(\mathbf{q}_{t+1}, \mathbf{h}_t | \Theta, \mathcal{X}_{t,t+1})$. The former distribution represents the error recursions (22) and (23) conditioned on the history of all preceding iterations, while the latter describes the error recursions (24) and (25). We follow the proof strategy in [12] to evaluate the two distributions via the conditional distribution $p(\mathbf{V} | \Theta, \mathcal{X}_{t,t'})$ for $t' = t$ or $t' = t + 1$. See Section I-B for the main idea in analyzing the conditional distributions.

Before analyzing the conditional distribution, we shall introduce several definitions used in the proof. For $t' > 0$, we partition $\mathbf{V} \in \mathcal{U}_N$ as

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{01}^{t,t'} \\ \mathbf{V}_{10}^{t,t'} \\ \mathbf{V}_{11}^{t,t'} \end{pmatrix}, \quad \mathbf{V} = \begin{pmatrix} \mathbf{V}_{00}^{t,t'} & \mathbf{V}_{01}^{t,t'} \\ \mathbf{V}_{10}^{t,t'} & \mathbf{V}_{11}^{t,t'} \end{pmatrix} \quad (60)$$

for $t > 0$, with $\mathbf{V}_{00}^{t,t'} \in \mathbb{C}^{t' \times t}$ and $\mathbf{V}_{11}^{t,t'} \in \mathbb{C}^{(N-t') \times (N-t)}$.

We write the SVDs of $\mathbf{V}_{10}^{t,t'} \in \mathbb{C}^{(N-t') \times t}$ and $\mathbf{M}_t \in \mathbb{C}^{N \times t}$ for $t > 0$ as

$$\mathbf{V}_{10}^{t,t'} = \Phi_{\mathbf{V}_{10}^{t,t'}} \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{V}_{10}^{t,t'}} \\ \mathbf{O} \end{pmatrix} \Psi_{\mathbf{V}_{10}^{t,t'}}^H, \quad (61)$$

$$\mathbf{M}_t = \Phi_{\mathbf{M}_t} \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{M}_t} \\ \mathbf{O} \end{pmatrix} \Psi_{\mathbf{M}_t}^H, \quad (62)$$

as defined in Section I-F. For notational convenience, we define $\Phi_{V_{10}^{0,t'}}^\perp = I_{N-t'}$ and $\Phi_{M_0}^\perp = I_N$. Similarly, we define the SVDs of $V_{01}^{t,t'} \in \mathbb{C}^{t' \times (N-t)}$ and $Q_{t'} \in \mathbb{C}^{N \times t'}$ for $t' > 0$ as

$$V_{01}^{t,t'} = \Phi_{V_{01}^{t,t'}} (\Sigma_{V_{01}^{t,t'}} O) \Psi_{V_{01}^{t,t'}}^H, \quad (63)$$

$$Q_{t'} = \Phi_{Q_{t'}} \begin{pmatrix} \Sigma_{Q_{t'}} \\ O \end{pmatrix} \Psi_{Q_{t'}}^H. \quad (64)$$

The following lemma provides a useful representation of $V \in \mathcal{U}_N$ conditioned on Θ and $\mathcal{X}_{t,t'}$, and corresponds to [12, Lemma 10].

Lemma 3: For $t \geq 0$, $t' > 0$, and $N - t - t' > 0$, suppose that $\tilde{V} \in \mathcal{U}_{N-t-t'}$ is a Haar matrix and independent of V , and that both $Q_{t'} \in \mathbb{C}^{N \times t'}$ and $M_t \in \mathbb{C}^{N \times t}$ are full rank for $t > 0$. Let

$$V_{00}^{t,t'} = (Q_{t'}^\dagger \Phi_{Q_{t'}}^\parallel)^H B_{t'}^H \Phi_{M_t}^\parallel \quad (65)$$

$$= (\Phi_{Q_{t'}}^\parallel)^H H_t M_t^\dagger \Phi_{M_t}^\parallel, \quad (66)$$

$$V_{01}^{t,t'} = (Q_{t'}^\dagger \Phi_{Q_{t'}}^\parallel)^H B_{t'}^H \Phi_{M_t}^\perp, \quad (67)$$

$$V_{10}^{t,t'} = (\Phi_{Q_{t'}}^\perp)^H H_t M_t^\dagger \Phi_{M_t}^\parallel, \quad (68)$$

with $V_{01}^{0,1} = b_0^H / \|q_0\|$. Then, the conditional distribution of V given Θ and $\mathcal{X}_{t,t'}$ for $t' = t$ or $t' = t + 1$ satisfies

$$V|_{\Theta, \mathcal{X}_{t,t'}} \sim \bar{V}_{t,t'} + \Phi_{Q_{t'}}^\perp \Phi_{V_{10}^{t,t'}}^\perp \tilde{V} (\Phi_{M_t}^\perp \Psi_{V_{01}^{t,t'}}^\perp)^H, \quad (69)$$

where the conditional mean $\bar{V}_{t,t'}$ is given by $\bar{V}_{0,1} = q_0 b_0^H / \|q_0\|^2$ for $t = 0$, and by

$$\bar{V}_{t,t'} = \Phi_{Q_{t'}} \begin{pmatrix} V_{00}^{t,t'} & V_{01}^{t,t'} \\ V_{10}^{t,t'} & \bar{V}_{11}^{t,t'} \end{pmatrix} \Phi_{M_t}^H \quad (70)$$

for $t > 0$, with

$$\bar{V}_{11}^{t,t'} = -V_{10}^{t,t'} [(V_{01}^{t,t'})^\dagger V_{00}^{t,t'}]^H \quad (71)$$

$$= -[V_{00}^{t,t'} (V_{10}^{t,t'})^\dagger]^H V_{01}^{t,t'}. \quad (72)$$

Proof: See Appendix D. ■

We next present a lemma to evaluate the conditional mean of V , which corresponds to [12, Lemma 12].

Lemma 4: For $t' \geq t > 0$ and $N - t - t' > 0$, suppose that both $Q_{t'} \in \mathbb{C}^{N \times t'}$ and $M_t \in \mathbb{C}^{N \times t}$ are full rank. Let

$$\epsilon_{1,t} = \Gamma_{t,t}^H H_t^H q_t^\perp, \quad (73)$$

$$\epsilon_{2,t} = \Delta_{t,t+1}^H B_{t+1}^H m_t^\perp, \quad (74)$$

with

$$\Gamma_{t,t'} = M_t^\dagger - M_t^\dagger B_{t'} (B_{t'}^H P_{M_t}^\perp B_{t'})^{-1} B_{t'}^H P_{M_t}^\perp, \quad (75)$$

$$\Delta_{t,t'} = Q_{t'}^\dagger - Q_{t'}^\dagger H_t (H_t^H P_{Q_{t'}}^\perp H_t)^{-1} H_t^H P_{Q_{t'}}^\perp. \quad (76)$$

Then, for all $\tau < t'$

$$\bar{V}_{t,t'}^H q_\tau = b_\tau, \quad (77)$$

$$\bar{V}_{t,t}^H q_t = B_t \beta_t + \epsilon_{1,t}, \quad (78)$$

$$\bar{V}_{t,t+1} m_t = H_t \alpha_t + \epsilon_{2,t}, \quad (79)$$

$$\Phi_{M_t}^\perp P_{V_{01}^{t,t'}}^\perp (\Phi_{M_t}^\perp)^H = P_{M_t}^\perp - P_{P_{M_t}^\perp B_{t'}}^\parallel, \quad (80)$$

$$\Phi_{Q_{t'}}^\perp P_{V_{10}^{t,t'}}^\perp (\Phi_{Q_{t'}}^\perp)^H = P_{Q_{t'}}^\perp - P_{P_{Q_{t'}}^\perp H_t}^\parallel. \quad (81)$$

Proof: See Appendix E. ■

The following lemma is used to prove the central limit theorem associated with the second term on the RHS of (69). Note that the lemma is not required for i.i.d. Gaussian measurement matrices, while evaluation of negligible terms is essentially the same as in [12, Lemma 2 (c)].

Lemma 5: For $t' > t \geq 0$ and $N - t - t' > 0$, suppose that $\tilde{V} \in \mathcal{U}_{N-t-t'}$ is a Haar matrix and independent of V . Let $a \in \mathbb{C}^{N-t-t'}$ denote a vector that are independent of \tilde{V} and satisfies $\lim_{N \rightarrow \infty} N^{-1} \|a\|^2 \stackrel{\text{a.s.}}{=} 1$. Suppose that $z \in \mathbb{C}^N$ is a vector such that, for any $k \in \mathbb{N}$, $z_N \sim \mathcal{CN}(0, I_k)$ holds for all subsets $\mathcal{N} \in \mathfrak{N}_k$ as $N \rightarrow \infty$.

- If the following properties hold:

$$\liminf_{M=\delta N \rightarrow \infty} \lambda_{\min} \left(\frac{1}{N} M_t^H M_t \right) \stackrel{\text{a.s.}}{>} C, \quad (82)$$

$$\liminf_{M=\delta N \rightarrow \infty} \lambda_{\min} \left(\frac{1}{N} B_t^H P_{M_t}^\perp B_t \right) \stackrel{\text{a.s.}}{>} C \quad (83)$$

for some constant $C > 0$, then

$$\Phi_{M_t}^\perp \Psi_{V_{01}^{t,t'}}^\perp \tilde{V}^H a \xrightarrow{d} z + M_t o(1) + P_{M_t}^\perp B_t o(1). \quad (84)$$

holds conditioned on a , Θ , and $\mathcal{X}_{t,t}$ in the large system limit.

- If the following properties hold:

$$\liminf_{M=\delta N \rightarrow \infty} \lambda_{\min} \left(\frac{1}{N} Q_{t+1}^H Q_{t+1} \right) \stackrel{\text{a.s.}}{>} C, \quad (85)$$

$$\liminf_{M=\delta N \rightarrow \infty} \lambda_{\min} \left(\frac{1}{N} H_t^H P_{Q_{t+1}}^\perp H_t \right) \stackrel{\text{a.s.}}{>} C \quad (86)$$

for some constant $C > 0$, then

$$\Phi_{Q_{t+1}}^\perp \Phi_{V_{10}^{t,t+1}}^\perp \tilde{V} a \xrightarrow{d} z + Q_{t+1} o(1) + P_{Q_{t+1}}^\perp H_t o(1) \quad (87)$$

holds conditioned on a , Θ , and $\mathcal{X}_{t,t+1}$ in the large system limit.

Proof: See Appendix F. ■

We are ready to prove Theorem 4. The proof is by induction. In the next subsection, we first prove the theorem for $\tau = 0$.

B. Proof for $\tau = 0$

Eqs. (49) for $\tau = 0$: The convergence (49) in distribution for $\tau = 0$ follows from (22), Theorem 2, $q_0 = x$, and Assumption 1. ■

Eqs. (51)–(54) for $\tau = 0$: The identity (51) for $\tau = 0$ follows from (22) and Lemma 1. Similarly, (22), Lemma 1, $q_0 = x$, and Assumption 1 imply that (52) holds for $\tau = 0$.

We next prove (53) for $\tau = 0$. Let $D = \tilde{W}_0(\Sigma, O)$, with \tilde{W}_0 given by (26). Assumption 2 implies that D satisfies the assumptions in Theorem 4. Thus, using (12), (51), and (52) for $\tau = 0$ yields

$$\lim_{M=\delta N \rightarrow \infty} \frac{\gamma_0}{N} b_0^H \tilde{W}_0 \{(\Sigma, O) b_0 + \tilde{w}\} \stackrel{\text{a.s.}}{=} \zeta_{0,0}. \quad (88)$$

From (23), (52) with $\mathbf{D} = \mathbf{I}_N$ for $\tau = 0$, and (88), we arrive at (53) for $\tau = 0$.

Finally, let us prove (54) for $\tau = 0$. Using (23) and (88), as well as (51) and (52) for $\tau = 0$, we have

$$\frac{1}{N} \mathbf{m}_0^H \mathbf{m}_0 \xrightarrow{\text{a.s.}} \frac{\gamma_0^2}{N} \mathbf{b}_0^H \mathbf{D} \mathbf{b}_0 + \frac{\gamma_0^2}{N} \tilde{\mathbf{w}}^H \tilde{\mathbf{W}}_0^H \tilde{\mathbf{W}}_0 \tilde{\mathbf{w}} - \zeta_{0,0} \quad (89)$$

in the large system limit, with

$$\mathbf{D} = \begin{pmatrix} \boldsymbol{\Sigma} \\ \mathbf{O} \end{pmatrix} \tilde{\mathbf{W}}_0^H \tilde{\mathbf{W}}_0 (\boldsymbol{\Sigma}, \mathbf{O}), \quad (90)$$

which satisfies the assumptions in Theorem 4, because of (26) and Assumption 2. Applying (52) for $\tau = 0$ and Assumption 3 to (89), we obtain (54) for $\tau = 0$. ■

Eq. (50) for $\tau = 0$: Applying Lemma 3 to (24) yields

$$\mathbf{h}_0 \sim \frac{\mathbf{b}_0^H \mathbf{m}_0}{\|\mathbf{q}_0\|^2} \mathbf{q}_0 + \boldsymbol{\Phi}_{\mathbf{Q}_1}^\perp \tilde{\mathbf{V}} (\boldsymbol{\Psi}_{\mathbf{V}_{01}^{0,1}}^\perp)^H \mathbf{m}_0, \quad (91)$$

conditioned on Θ and $\mathcal{X}_{0,1}$. From (53) for $\tau = 0$, we find that the coefficient $\mathbf{b}_0^H \mathbf{m}_0 / \|\mathbf{q}_0\|^2$ in the first term converges almost surely to zero in the large system limit.

We define the variance ν_0 in (46) as

$$\nu_0 \stackrel{\text{a.s.}}{=} \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{m}_0^H \mathbf{P}_{\mathbf{V}_{01}^{0,1}}^\perp \mathbf{m}_0. \quad (92)$$

Lemma 5 implies that, for any $k \in \mathbb{N}$, $\mathbf{h}_{0,\mathcal{N}}$ given by (91) converges in distribution to $\tilde{\mathbf{h}}_{0,\mathcal{N}}$ given by (46) for all $\mathcal{N} \in \mathfrak{N}_k$ in the large system limit, when ν_0 is defined by (92).

In order to complete the proof, we shall prove that (92) reduces to (48). Applying $\mathbf{V}_{01}^{0,1} = \mathbf{b}_0^H / \|\mathbf{q}_0\|$ and $\mathbf{P}_{\mathbf{V}_{01}^{0,1}}^\perp = \mathbf{I}_N - (\mathbf{V}_{01}^{0,1})^H \{ \mathbf{V}_{01}^{0,1} (\mathbf{V}_{01}^{0,1})^H \}^{-1} \mathbf{V}_{01}^{0,1}$ to (92), and using $\mathbf{m}_0^\perp = \mathbf{m}_0$, (52), and (53) for $\tau = 0$, we have (48) for $\tau = 0$. Thus, (50) holds for $\tau = 0$. ■

Eq. (55) for $\tau = 0$: Let us prove (55) for $\tau = 0$. Using (91) and Lemma 1 yields

$$\frac{1}{N} \mathbf{h}_0^H \mathbf{h}_0 \xrightarrow{\text{a.s.}} \frac{1}{N} \frac{|\mathbf{b}_0^H \mathbf{m}_0|^2}{\|\mathbf{q}_0\|^2} + \nu_0 \quad (93)$$

conditioned on Θ and $\mathcal{X}_{0,1}$ in the large system limit. From (53) for $\tau = 0$, we find that the first term vanishes in the large system limit. Since ν_0 given by (48) is equal to the RHS of (55) for $\tau = 0$ and $\tau' = 0$, (55) holds for $\tau = 0$. ■

Eq. (41) for $\tau = 0$: We shall prove (41) for $\tau = 0$. For any $a \geq 0$, $b \geq 0$, and $a + b \leq c$, we utilize the fact that, for random variables X and Y , $\mathbb{E}[|X|^a |Y|^b]$ is bounded if both $\mathbb{E}[|X|^c]$ and $\mathbb{E}[|Y|^c]$ are bounded. The fact is trivial for $a = 0$ or $b = 0$. Otherwise, using Hölder's inequality yields

$$\mathbb{E}[|X|^a |Y|^b] \leq (\mathbb{E}[|X|^c])^{a/c} (\mathbb{E}[|Y|^c])^{b/c}, \quad (94)$$

with $q = c/(c - a) > 0$. Since the condition $a + b \leq c$ implies $bq \leq c$, we find $\mathbb{E}[|X|^a |Y|^b] < \infty$. Thus, from (25) and Assumption 1 it is sufficient to prove that $\mathbb{E}[|\eta_0(q_{0,n} - h_{0,n})|^4] < \infty$ holds in the large system limit.

Using (50) for $\tau = 0$, we have

$$\mathbb{E}[|\eta_0(q_{0,n} - h_{0,n})|^4] \rightarrow \mathbb{E}[|\eta_0(\tilde{x}_{n,A \rightarrow B}^0)|^4] \quad (95)$$

in the large system limit, with $\tilde{x}_{n,A \rightarrow B}^0 = q_{0,0} - \nu_0^{1/2} z_{0,0}$. From (19), the moment (95) is bounded if $\mathbb{E}[|\tilde{\eta}_0(\tilde{x}_{n,A \rightarrow B}^0)|^4]$

is bounded. From (54) for $\tau = 0$, the variance ν_0 given in (48) coincides with $v_{A \rightarrow B}^0$ given by (10). Furthermore, the MMSE estimator $\tilde{\eta}_0(\tilde{x}_{n,A \rightarrow B}^0)$ is equal to the posterior mean $\mathbb{E}[x_n | x_{n,A \rightarrow B}^0 = \tilde{x}_{n,A \rightarrow B}^0]$ of x_n defined via (16). From these observations, we use Jensen's inequality to obtain

$$\begin{aligned} \mathbb{E}[|\tilde{\eta}_0(\tilde{x}_{n,A \rightarrow B}^0)|^4] &\leq \mathbb{E}[\mathbb{E}[|x_n|^4 | x_{n,A \rightarrow B}^0 = \tilde{x}_{n,A \rightarrow B}^0]] \\ &= \mathbb{E}[|x_n|^4] < \infty, \end{aligned} \quad (96)$$

because of Assumption 1. Thus, (41) holds for $\tau = 0$. ■

Eq. (42) for $\tau = 0$: We shall prove the existence of the limit (42) for $\tau = 0$. We only consider the case $\tau' = 0$, since the case $\tau' = 1$ can be proved in the same manner. Using (25) yields

$$\frac{1}{N} \mathbf{q}_0^H \mathbf{q}_1 = \frac{\|\mathbf{q}_0\|^2}{N} - \frac{\mathbf{q}_0^H \eta_0(\mathbf{q}_0 - \mathbf{h}_0)}{N}. \quad (97)$$

Since the first term converges almost surely to 1 in the large system limit, it is sufficient to prove that the second term is convergent in the large system limit.

In order to utilize Theorem 1, we use (50) for $\tau = 0$ to note

$$\begin{aligned} \frac{1}{N} \mathbb{V}[\mathbf{q}_0^H \eta_0(\mathbf{q}_0 - \mathbf{h}_0)] &\rightarrow \frac{1}{N} \mathbb{V}[\mathbf{q}_0^H \eta_0(\mathbf{q}_0 - \tilde{\mathbf{h}}_0)] \\ &= \mathbb{V}[\mathbf{q}_{0,0}^* \eta_0(\tilde{x}_{0,A \rightarrow B}^0)] \end{aligned} \quad (98)$$

in the large system limit. We repeat the proof of (41) to find the boundedness of (98). Thus, we can use Theorem 1 to find

$$\begin{aligned} \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{q}_0^H \eta_0(\mathbf{q}_0 - \mathbf{h}_0) \\ \stackrel{\text{a.s.}}{=} \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbb{E}[\mathbf{q}_0^H \eta_0(\mathbf{q}_0 - \tilde{\mathbf{h}}_0)] < \infty. \end{aligned} \quad (99)$$

Thus, the limit (42) exists for $\tau' = 0$ and $\tau = 0$. ■

Eq. (56) for $\tau = 0$: We evaluate $N^{-1} \mathbf{h}_0^H \mathbf{q}_{\tau''}$. For $\tau'' = 0$, from (91) we have

$$\frac{1}{N} \mathbf{h}_0^H \mathbf{q}_0 \sim \frac{1}{N} \mathbf{m}_0^H \mathbf{b}_0 \quad (100)$$

conditioned on Θ and $\mathcal{X}_{0,1}$, where we have used $(\boldsymbol{\Phi}_{\mathbf{Q}_1}^\perp)^H \mathbf{q}_0 = \mathbf{0}$. From (53) for $\tau = 0$, we find that $N^{-1} \mathbf{h}_0^H \mathbf{q}_0$ converges almost surely to zero in the large system limit.

We next evaluate $N^{-1} \mathbf{h}_0^H \mathbf{q}_1$. Using (25) yields

$$\frac{1}{N} \mathbf{h}_0^H \mathbf{q}_1 = \frac{1}{N} \mathbf{h}_0^H \mathbf{q}_0 - \frac{1}{N} \mathbf{h}_0^H \eta_0(\mathbf{q}_0 - \mathbf{h}_0) \quad (101)$$

conditioned on Θ and $\mathcal{X}_{0,1}$, where \mathbf{h}_0 is given by (91). We have proved that the first term converges almost surely to zero in the large system limit. Repeating the proof of (42) for $\tau = 0$, we use Theorem 1 to find that (101) reduces to

$$\frac{1}{N} \mathbf{h}_0^H \mathbf{q}_1 \xrightarrow{\text{a.s.}} -\frac{1}{N} \mathbb{E}[\tilde{\mathbf{h}}_0^H \eta_0(\mathbf{q}_0 - \tilde{\mathbf{h}}_0)] \quad (102)$$

in the large system limit.

In order to evaluate (102), we utilize Lemma 2. Since the variance ν_0 in (46) is equal to $v_{A \rightarrow B}^0$ given by (10), we can use Lemma 2 to obtain

$$\frac{1}{N} \mathbf{h}_0^H \mathbf{q}_1 \xrightarrow{\text{a.s.}} \frac{o(1)}{N} \mathbb{E}[\mathbf{q}_0^H \eta_0(\mathbf{q}_0 - \tilde{\mathbf{h}}_0)] \xrightarrow{\text{a.s.}} 0 \quad (103)$$

in the large system limit. Thus, (56) holds for $\tau = 0$. ■

Eqs. (58) and (59) for $\tau = 0$: We use (50) for $\tau = 0$ to find that the individual MSE (33) reduces to

$$\text{mse}_n^0 = \lim_{M=\delta N \rightarrow \infty} \mathbb{E} \left[|q_{0,n} - \tilde{\eta}_0(q_{0,n} - \tilde{h}_{0,n})|^2 \right], \quad (104)$$

which is equal to $\text{MMSE}(v_{A \rightarrow B}^0)$ given by (15), because of $\nu_0 = v_{A \rightarrow B}^0$ in (46). Thus, (58) holds for $\tau = 0$.

Let us prove (59) for $\tau = 0$. Applying (19) to (25) for $\tau = 0$, and using (18), we have

$$q_{1,n} = \frac{v_{B \rightarrow A}^1 \{q_{0,n} - \tilde{\eta}_0(q_{0,n} - h_{0,n})\}}{\text{MMSE}(v_{A \rightarrow B}^0)} - \frac{v_{B \rightarrow A}^1}{v_{A \rightarrow B}^0} h_{0,n}. \quad (105)$$

In the same manner as in the proof of (58), evaluating (32) with (105) yields

$$\text{mse}_{n,B \rightarrow A}^1 \stackrel{\text{a.s.}}{=} \frac{(v_{B \rightarrow A}^1)^2}{\text{MMSE}(v_{A \rightarrow B}^0)} - \frac{(v_{B \rightarrow A}^1)^2}{v_{A \rightarrow B}^0}, \quad (106)$$

where we have used $\mathbb{E}[h_{0,n}^* q_{0,n}] = o(1) \mathbb{E}[|q_{0,n}|^2] \rightarrow 0$ in the large system limit, obtained from (46) and (50) for $\tau = 0$, as well as Lemma 2 and (58) for $\tau = 0$. From (18) and (106), we arrive at (59) for $\tau = 0$. ■

Eqs. (43) and (44) for $\tau = 0$: We only prove (44) for $\tau = 0$, since (43) is trivial. If $\liminf_{M=\delta N \rightarrow \infty} N^{-1} \|\mathbf{q}_1^\perp\|^2$ converges almost surely to a strictly positive constant, (44) holds for $\tau = 0$ [12, Lemmas 8 and 9]. By definition, we have

$$\|\mathbf{q}_1^\perp\|^2 = \mathbf{q}_1^H \mathbf{P}_{\mathbf{Q}_1}^\perp \mathbf{q}_1 = \eta_0^H(\mathbf{q}_0 - \mathbf{h}_0) \mathbf{P}_{\mathbf{Q}_1}^\perp \eta_0(\mathbf{q}_0 - \mathbf{h}_0), \quad (107)$$

where (25) has been used. Repeating the proof of (42) for $\tau = 0$, we find

$$\begin{aligned} \frac{1}{N} \|\mathbf{q}_1^\perp\|^2 &\stackrel{\text{a.s.}}{\rightarrow} \mathbb{E} \left[|\eta_0(\tilde{x}_{0,A \rightarrow B}^0)|^2 \right] \\ &- \frac{1}{N} \sum_{n,n'} \mathbb{E} \left[\eta_0^*(\tilde{x}_{n,A \rightarrow B}^0) [P_{\mathbf{Q}_1}^\parallel]_{nn'} \eta_0(\tilde{x}_{n',A \rightarrow B}^0) \right] \end{aligned} \quad (108)$$

in the large system limit.

We next show that the second term on the RHS of (108) reduces to

$$\begin{aligned} &\frac{1}{N} \sum_{n,n'} \mathbb{E} \left[\eta_0^*(\tilde{x}_{n,A \rightarrow B}^0) [P_{\mathbf{Q}_1}^\parallel]_{nn'} \eta_0(\tilde{x}_{n',A \rightarrow B}^0) \right] \\ &\rightarrow \frac{1}{N} \mathbb{E} \left\{ \mathbb{E}_{z_0} [\eta_0^H(\tilde{x}_{A \rightarrow B}^0)] P_{\mathbf{Q}_1}^\parallel \mathbb{E}_{z_0} [\eta_0(\tilde{x}_{A \rightarrow B}^0)] \right\} \end{aligned} \quad (109)$$

in the large system limit, with $[\tilde{x}_{A \rightarrow B}^0]_n = \tilde{x}_{n,A \rightarrow B}^0$. Applying the Cauchy-Schwarz inequality to the terms with $n = n'$, we have

$$\begin{aligned} &\left(\frac{1}{N} \sum_{n=0}^{N-1} [P_{\mathbf{Q}_1}^\parallel]_{nn} |\eta_0(\tilde{x}_{n,A \rightarrow B}^0)|^2 \right)^2 \\ &\stackrel{\text{a.s.}}{\leq} \frac{\mathbb{E}[|\eta_0(\tilde{x}_{0,A \rightarrow B}^0)|^4]}{N} \sum_{n=0}^{N-1} [P_{\mathbf{Q}_1}^\parallel]_{nn}^2 \stackrel{\text{a.s.}}{\rightarrow} 0 \end{aligned} \quad (110)$$

in the large system limit. The convergence to zero follows from the boundedness of $\mathbb{E}[|\eta_0(\tilde{x}_{0,A \rightarrow B}^0)|^4]$ and the upper bound $\sum_{n=0}^{N-1} [P_{\mathbf{Q}_1}^\parallel]_{nn}^2 < \text{Tr}\{(\mathbf{P}_{\mathbf{Q}_1}^\parallel)^2\} = \text{Tr}(\mathbf{P}_{\mathbf{Q}_1}^\parallel) = 1$. Similarly, we find

$$\left(\frac{1}{N} \sum_{n=0}^{N-1} [P_{\mathbf{Q}_1}^\parallel]_{nn} |\mathbb{E}_{z_0,n} [\eta_0(\tilde{x}_{n,A \rightarrow B}^0)]|^2 \right)^2 \stackrel{\text{a.s.}}{\rightarrow} 0 \quad (111)$$

in the large system limit. Thus, (109) holds.

In order to bound (108) with (109), we use the fact that the maximum eigenvalue of the projection matrix $\mathbf{P}_{\mathbf{Q}_1}^\parallel$ is 1.

$$\begin{aligned} \liminf_{M=\delta N \rightarrow \infty} \frac{1}{N} \|\mathbf{q}_1^\perp\|^2 &\stackrel{\text{a.s.}}{\geq} \mathbb{E} \left[|\eta_0(\tilde{x}_{0,A \rightarrow B}^0)|^2 \right] \\ &- \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mathbb{E} \left[|\mathbb{E}_{z_0,n} [\eta_0(\tilde{x}_{n,A \rightarrow B}^0)]|^2 \right] \\ &= \mathbb{E} \left\{ \mathbb{V}_{z_0,0} [\eta_0(\tilde{x}_{0,A \rightarrow B}^0)] \right\} \end{aligned} \quad (112)$$

in the large system limit. Since $\eta_0(\tilde{x}_{0,A \rightarrow B}^0)$ is a non-constant random variable, the lower bound (112) is strictly positive. Thus, (44) holds for $\tau = 0$. ■

C. Proof by Induction

We have proved that Theorem 4 holds for $\tau = 0$. Next, we assume that Theorem 4 is correct for all $\tau < t$, and prove that Theorem 4 holds for $\tau = t$. Note that we can use Lemmas 3 and 4, since the induction hypotheses (43) and (44) $\tau < t$ imply that \mathbf{M}_t and $\mathbf{Q}_{t'}$ are full rank for $t' = t$ and $t' = t+1$.

Eq. (49) for $\tau = t$: Using (22), Lemma 3, and Lemma 4 yields

$$\mathbf{b}_t \sim \mathbf{B}_t \boldsymbol{\beta}_t + \boldsymbol{\epsilon}_{1,t} + \boldsymbol{\Phi}_{\mathbf{M}_t}^\perp \boldsymbol{\Psi}_{\mathbf{V}_{01}^{t,t}}^\perp \tilde{\mathbf{V}}^H (\boldsymbol{\Phi}_{\mathbf{Q}_t}^\perp \boldsymbol{\Phi}_{\mathbf{V}_{10}^{t,t}}^\perp)^H \mathbf{q}_t \quad (113)$$

conditioned on Θ and $\mathcal{X}_{t,t}$, with $\boldsymbol{\epsilon}_{1,t}$ given by (73).

Let us prove $\boldsymbol{\epsilon}_{1,t} \stackrel{\text{a.s.}}{=} o(1)$ in the large system limit. We use the submultiplicative property of the Euclidean norm to obtain

$$\|\boldsymbol{\epsilon}_{1,t}\|^2 \leq \|N \boldsymbol{\Gamma}_{t,t}\|^2 \|N^{-1} \mathbf{H}_t^H \mathbf{q}_t^\perp\|^2. \quad (114)$$

We first prove that $N^{-1} \mathbf{H}_t^H \mathbf{q}_t^\perp$ converges almost surely to zero in the large system limit. By definition,

$$\frac{1}{N} \mathbf{H}_t^H \mathbf{q}_t^\perp = \frac{\mathbf{H}_t^H \mathbf{q}_t}{N} - \frac{\mathbf{H}_t^H \mathbf{Q}_t}{N} \left(\frac{\mathbf{Q}_t^H \mathbf{Q}_t}{N} \right)^{-1} \frac{\mathbf{Q}_t^H \mathbf{q}_t}{N}. \quad (115)$$

The induction hypothesis (56) for $\tau < t$ implies that $N^{-1} \mathbf{H}_t^H \mathbf{q}_t$ and $N^{-1} \mathbf{H}_t^H \mathbf{Q}_t$ converge almost surely to zero in the large system limit. Furthermore, the induction hypotheses (42) and (44) for $\tau < t$ imply that $\|(N^{-1} \mathbf{Q}_t^H \mathbf{Q}_t)^{-1} N^{-1} \mathbf{Q}_t^H \mathbf{q}_t\|$ is bounded. Thus, $N^{-1} \mathbf{H}_t^H \mathbf{q}_t^\perp$ converges almost surely to zero in the large system limit.

In order to complete the proof of $\boldsymbol{\epsilon}_{1,t} \stackrel{\text{a.s.}}{=} o(1)$, we shall prove that $\|N \boldsymbol{\Gamma}_{t,t}\|^2$ is bounded. Using (75) and $\mathbf{M}_t^\dagger \mathbf{P}_{\mathbf{M}_t}^\perp = \mathbf{O}$ yields

$$\begin{aligned} \|N \boldsymbol{\Gamma}_{t,t}\|^2 &= \text{Tr} \left\{ \left(\frac{\mathbf{M}_t^H \mathbf{M}_t}{N} \right)^{-1} + \left(\frac{\mathbf{B}_t^H \mathbf{P}_{\mathbf{M}_t}^\perp \mathbf{B}_t}{N} \right)^{-1} \right. \\ &\quad \left. \cdot \frac{\mathbf{B}_t^H \mathbf{M}_t}{N} \left(\frac{\mathbf{M}_t^H \mathbf{M}_t}{N} \right)^{-2} \frac{\mathbf{M}_t^H \mathbf{B}_t}{N} \right\}. \end{aligned} \quad (116)$$

The induction hypothesis (43) for $\tau < t$ implies that the first term into the trace is bounded in the large system limit. Furthermore, we use the induction hypotheses (43), (52), and (53) for $\tau < t$ to obtain

$$\lim_{M=\delta N \rightarrow \infty} \left(\frac{\mathbf{B}_t^H \mathbf{P}_{\mathbf{M}_t}^\perp \mathbf{B}_t}{N} \right)^{-1} \stackrel{\text{a.s.}}{=} \lim_{M=\delta N \rightarrow \infty} \left(\frac{\mathbf{Q}_t^H \mathbf{Q}_t}{N} \right)^{-1}, \quad (117)$$

which implies the boundedness of $\|(N^{-1}\mathbf{B}_t^\mathbf{H}\mathbf{P}_{\mathbf{M}_t}^\perp\mathbf{B}_t)^{-1}\|$, because of the induction hypothesis (44) for $\tau < t$. Thus, $\|\mathbf{N}\mathbf{T}_{t,t}\|^2$ is bounded in the large system limit.

We have proved $\epsilon_{1,t} = o(1)$. We next use Lemma 5 to evaluate the second term on the RHS of (113). It is possible to confirm that the last term on the RHS of (84) reduces to

$$\mathbf{P}_{\mathbf{M}_t}^\perp\mathbf{B}_t o(1) = \mathbf{M}_t o(1) + \mathbf{B}_t o(1). \quad (118)$$

Thus, we use (113) and Lemma 5 to find that (49) holds for $\tau = t$, when μ_t in (45) is defined as

$$\mu_t \stackrel{\text{a.s.}}{=} \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{q}_t^\mathbf{H} \Phi_{\mathbf{Q}_t}^\perp \mathbf{P}_{\mathbf{V}_{10}^{t,t}}^\perp (\Phi_{\mathbf{Q}_t}^\perp)^\mathbf{H} \mathbf{q}_t. \quad (119)$$

In order to complete the proof of (49), we shall prove that (119) reduces to (47). From Lemma 4, it is sufficient to prove

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{q}_t^\mathbf{H} \mathbf{P}_{\mathbf{Q}_t}^\perp \mathbf{H}_t \mathbf{q}_t \stackrel{\text{a.s.}}{=} 0. \quad (120)$$

By definition, we have

$$\frac{\mathbf{q}_t^\mathbf{H} \mathbf{P}_{\mathbf{Q}_t}^\perp \mathbf{H}_t \mathbf{q}_t}{N} = \left(\frac{\mathbf{H}_t^\mathbf{H} \mathbf{q}_t^\perp}{N} \right)^\mathbf{H} \left(\frac{\mathbf{H}_t^\mathbf{H} \mathbf{P}_{\mathbf{Q}_t}^\perp \mathbf{H}_t}{N} \right)^{-1} \frac{\mathbf{H}_t^\mathbf{H} \mathbf{q}_t^\perp}{N}. \quad (121)$$

We have already proved $N^{-1} \mathbf{H}_t^\mathbf{H} \mathbf{q}_t^\perp \xrightarrow{\text{a.s.}} 0$ in the large system limit. Using the induction hypotheses (44), (55), and (56) for $\tau < t$ to evaluate $\mathbf{H}_t^\mathbf{H} \mathbf{P}_{\mathbf{Q}_t}^\perp \mathbf{H}_t$ yields

$$\lim_{M=\delta N \rightarrow \infty} \left(\frac{\mathbf{H}_t^\mathbf{H} \mathbf{P}_{\mathbf{Q}_t}^\perp \mathbf{H}_t}{N} \right)^{-1} \stackrel{\text{a.s.}}{\rightarrow} \lim_{M=\delta N \rightarrow \infty} \left(\frac{\mathbf{M}_t^\mathbf{H} \mathbf{M}_t}{N} \right)^{-1}, \quad (122)$$

which implies the boundedness of $\|(N^{-1} \mathbf{H}_t^\mathbf{H} \mathbf{P}_{\mathbf{Q}_t}^\perp \mathbf{H}_t)^{-1}\|$ in the large system limit, because of the induction hypothesis (43) for $\tau < t$. From these observations, we have (120). Thus, (49) holds for $\tau = t$. ■

Eqs. (51)–(54) for $\tau = t$: We first prove (52) for $\tau = t$. We use (113), $\epsilon_{1,t} = o(1)$, and Lemma 1 to have

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{b}_{\tau'}^\mathbf{H} \mathbf{D} \mathbf{b}_t \stackrel{\text{a.s.}}{=} \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{b}_{\tau'}^\mathbf{H} \mathbf{D} \mathbf{B}_t \beta_t \quad (123)$$

conditioned on Θ and $\mathcal{X}_{t,t}$ for $\tau < t$. Using the induction hypothesis (52) for $\tau < t$, $\mathbf{q}_t^\parallel = \mathbf{Q}_t \beta_t$, and $\mathbf{q}_{\tau'}^\mathbf{H} \mathbf{q}_t^\perp = 0$ yields (52) for $\tau' < \tau = t$.

For $\tau' = t$, (113) and Lemma 1 imply

$$\begin{aligned} \frac{1}{N} \mathbf{b}_t^\mathbf{H} \mathbf{D} \mathbf{b}_t &\stackrel{\text{a.s.}}{\rightarrow} \frac{1}{N} \beta_t^\mathbf{H} \mathbf{B}_t^\mathbf{H} \mathbf{D} \mathbf{B}_t \beta_t \\ &+ \frac{\mu_t}{N} \text{Tr} \left\{ \mathbf{D} \Phi_{\mathbf{M}_t}^\perp \mathbf{P}_{\mathbf{V}_{01}^{t,t}}^\perp (\Phi_{\mathbf{M}_t}^\perp)^\mathbf{H} \right\} \end{aligned} \quad (124)$$

conditioned on Θ and $\mathcal{X}_{t,t}$ in the large system limit, with μ_t given by (47). The induction hypothesis (52) for $\tau < t$ implies that the first term converges almost surely to $\lim_{M=\delta N \rightarrow \infty} N^{-1} \|\mathbf{q}_t^\parallel\|^2 N^{-1} \text{Tr}(\mathbf{D})$. Thus, we complete the proof of (52) for $\tau' = \tau = t$, by proving

$$\frac{1}{N} \text{Tr} \left\{ \mathbf{D} \Phi_{\mathbf{M}_t}^\perp \mathbf{P}_{\mathbf{V}_{01}^{t,t}}^\perp (\Phi_{\mathbf{M}_t}^\perp)^\mathbf{H} \right\} \stackrel{\text{a.s.}}{\rightarrow} \frac{1}{N} \text{Tr}(\mathbf{D}) \quad (125)$$

in the large system limit. From (80), it is sufficient to prove

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \text{Tr}(\mathbf{D} \mathbf{P}_{\mathbf{M}_t}^\perp) \stackrel{\text{a.s.}}{=} 0, \quad (126)$$

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \text{Tr}(\mathbf{D} \mathbf{P}_{\mathbf{M}_t}^\perp \mathbf{B}_t) \stackrel{\text{a.s.}}{=} 0. \quad (127)$$

We only prove (127), since (126) can be proved in the same manner. Using the Cauchy-Schwarz inequality yields

$$\left\{ \frac{1}{N} \text{Tr} \left(\mathbf{D} \mathbf{P}_{\mathbf{M}_t}^\perp \mathbf{B}_t \right) \right\}^2 \leq \frac{\|\mathbf{D}\|^2}{N} \frac{\|\mathbf{P}_{\mathbf{M}_t}^\perp \mathbf{B}_t\|^2}{N}. \quad (128)$$

From the assumptions in Theorem 4, $N^{-1} \|\mathbf{D}\|^2$ is bounded as $N \rightarrow \infty$. Furthermore, $\|\mathbf{P}_{\mathbf{M}_t}^\perp \mathbf{B}_t\|^2 = t$ holds. From these observations, we arrive at (127). Thus, (52) holds.

In order to prove (51) for $\tau = t$, we repeat the same argument to obtain

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{b}_t^\mathbf{H} \boldsymbol{\omega} \stackrel{\text{a.s.}}{=} \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \beta_t^\mathbf{H} \mathbf{B}_t^\mathbf{H} \boldsymbol{\omega} \stackrel{\text{a.s.}}{=} 0 \quad (129)$$

conditioned on Θ and $\mathcal{X}_{t,t}$, where we have used the induction hypothesis (51) for $\tau < t$.

Let us prove (53) and (54) for $\tau = t$. Using (12), (26), (51), and (52), we obtain

$$\lim_{M=\delta N \rightarrow \infty} \frac{\gamma_t}{N} \mathbf{b}_{\tau'}^\mathbf{H} \tilde{\mathbf{W}}_t \{(\boldsymbol{\Sigma}, \mathbf{O}) \mathbf{b}_t + \tilde{\mathbf{w}}\} \stackrel{\text{a.s.}}{=} \zeta_{t,\tau'}. \quad (130)$$

The property (53) follows from (23) and (130).

Similarly, we use (23), (52), and (130) to obtain

$$\begin{aligned} \frac{\mathbf{m}_{\tau'}^\mathbf{H} \mathbf{m}_t}{N} &\stackrel{\text{a.s.}}{\rightarrow} -\zeta_{t,\tau'} + \frac{\gamma_{\tau'} \gamma_t}{N} \{(\boldsymbol{\Sigma}, \mathbf{O}) \mathbf{b}_{\tau'} + \tilde{\mathbf{w}}\}^\mathbf{H} \\ &\quad \cdot \tilde{\mathbf{W}}_{\tau'}^\mathbf{H} \tilde{\mathbf{W}}_t \{(\boldsymbol{\Sigma}, \mathbf{O}) \mathbf{b}_t + \tilde{\mathbf{w}}\} \end{aligned} \quad (131)$$

in the large system limit. Using (26), (51), (52), and Assumption 2, we find that the second term reduces to (57) for $t' = \tau'$. Thus, (54) holds for $\tau = t$. ■

Eq. (50) for $\tau = t$: We shall prove (50) for $\tau = t$. Using (24) and Lemma 3 yields

$$\mathbf{h}_t \sim \bar{\mathbf{V}}_{t,t+1} \mathbf{m}_t + \Phi_{\mathbf{Q}_{t+1}}^\perp \Phi_{\mathbf{V}_{10}^{t,t+1}}^\perp \tilde{\mathbf{V}} (\Phi_{\mathbf{M}_t}^\perp \Psi_{\mathbf{V}_{01}^{t,t+1}}^\perp)^\mathbf{H} \mathbf{m}_t \quad (132)$$

conditioned on Θ and $\mathcal{X}_{t,t+1}$. Applying Lemma 4, we have

$$\mathbf{h}_t \sim \mathbf{H}_t \boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_{2,t} + \Phi_{\mathbf{Q}_{t+1}}^\perp \Phi_{\mathbf{V}_{10}^{t,t+1}}^\perp \tilde{\mathbf{V}} (\Phi_{\mathbf{M}_t}^\perp \Psi_{\mathbf{V}_{01}^{t,t+1}}^\perp)^\mathbf{H} \mathbf{m}_t \quad (133)$$

conditioned on Θ and $\mathcal{X}_{t,t+1}$, where $\boldsymbol{\epsilon}_{2,t}$ is given by (74). It is possible to prove $\boldsymbol{\epsilon}_{2,t} = o(1)$ in the large system limit, by repeating the proof of $\epsilon_{1,t} = o(1)$.

Lemma 5 implies that $\mathbf{h}_{t,\mathcal{N}}$ conditioned on Θ and $\mathcal{X}_{t,t+1}$ converges in distribution to $\tilde{\mathbf{h}}_{t,\mathcal{N}}$ given by (46) in the large system limit, when ν_t is defined as

$$\nu_t = \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{m}_t^\mathbf{H} \Phi_{\mathbf{M}_t}^\perp \mathbf{P}_{\mathbf{V}_{01}^{t,t+1}}^\perp (\Phi_{\mathbf{M}_t}^\perp)^\mathbf{H} \mathbf{m}_t. \quad (134)$$

It is possible to prove that (134) reduces to (48), by repeating the evaluate of (119). Thus, (50) holds for $\tau = t$. ■

Eq. (55) for $\tau = t$: We next prove (55) for $\tau = t$. From (133), $\boldsymbol{\epsilon}_{2,t} = o(1)$, and Lemma 1, we have

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{h}_{\tau'}^\mathbf{H} \mathbf{h}_t \stackrel{\text{a.s.}}{=} \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{h}_{\tau'}^\mathbf{H} \mathbf{H}_t \boldsymbol{\alpha}_t \quad (135)$$

conditioned on Θ and $\mathcal{X}_{t,t+1}$ for $\tau' < t$. Using the induction hypothesis (55) for $\tau < t$, $\mathbf{m}_t^\parallel = \mathbf{M}_t \boldsymbol{\alpha}_t$, and $\mathbf{m}_{\tau'}^\mathbf{H} \mathbf{m}_t^\perp = 0$ yields (55) for $\tau' < \tau = t$.

For $\tau' = t$, we use (133) and Lemma 1 to have

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{h}_t^H \mathbf{h}_t \stackrel{\text{a.s.}}{=} \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \boldsymbol{\alpha}_t^H \mathbf{H}_t^H \mathbf{H}_t \boldsymbol{\alpha}_t + \nu_t \quad (136)$$

conditioned on Θ and $\mathcal{X}_{t,t+1}$ in the large system limit, with ν_t given by (48). Applying the induction hypothesis (55) for $\tau < t$ and $\mathbf{m}_t^\parallel = \mathbf{M}_t \boldsymbol{\alpha}_t$ to the first term, we arrive at (55) for $\tau = \tau' = t$. Thus, (55) holds for $\tau = t$. ■

Eq. (41) for $\tau = t$: Let us prove (41) for $\tau = t$. It is sufficient to prove that $\mathbb{E}[|\tilde{\eta}_t(q_{0,n} - h_{t,n})|^4] < \infty$ holds in the large system limit. We first confirm that

$$\mathbb{E}[|\tilde{\eta}_t(q_{0,n} - h_{t,n})|^4] \stackrel{\text{a.s.}}{\rightarrow} \mathbb{E}[|\tilde{\eta}_t(\tilde{x}_{n,A \rightarrow B}^t)|^4] \quad (137)$$

holds in the large system limit, with $\tilde{x}_{n,A \rightarrow B}^t = q_{0,n} - h_{t,n}^G$, in which $h_{t,n}^G$ is the n th element of \mathbf{h}_t^G , recursively defined as

$$\mathbf{h}_\tau^G = \mathbf{H}_\tau^G \boldsymbol{\alpha}_\tau + \mathbf{Q}_{\tau+1} \mathbf{o}(1) + \mathbf{H}_\tau \mathbf{o}(1) + \nu_\tau^{1/2} \mathbf{z}_\tau, \quad (138)$$

with $\mathbf{H}_\tau^G = (\mathbf{h}_0^G, \dots, \mathbf{h}_{\tau-1}^G)$. Using (50) for $\tau = t$ yields

$$\mathbb{E}[|\tilde{\eta}_t(q_{0,n} - h_{t,n})|^4] \stackrel{\text{a.s.}}{\rightarrow} \mathbb{E}[|\tilde{\eta}_t(q_{0,n} - \tilde{h}_{t,n})|^4] \quad (139)$$

in the large system limit. Applying the induction hypothesis (50) repeatedly in the order $\tau = t-1, \dots, 0$, we arrive at (137).

We next evaluate the RHS of (137). From (54), (55), and (57) for $\tau = \tau' = t$, as well as from the induction hypothesis (59) for $\tau = t-1$, we find that the random vector \mathbf{h}_t^G induced from the randomness of $\mathcal{Z}_t = \{\mathbf{z}_\tau : \tau = 0, \dots, t\}$ has i.i.d. proper complex Gaussian elements with vanishing mean in the large system limit and variance $v_{A \rightarrow B}^t$. Repeating the proof of (41) for $\tau = 0$, we obtain

$$\lim_{M=\delta N \rightarrow \infty} \mathbb{E}[|\tilde{\eta}_t(\tilde{x}_{n,A \rightarrow B}^t)|^4] \leq \mathbb{E}[|x_n|^4] < \infty. \quad (140)$$

Thus, (41) holds for $\tau = t$. ■

Eq. (42) for $\tau = t$: We only prove the existence of (42) for $\tau' \leq \tau = t$, since the case $\tau' = t+1$ can be proved in the same manner. Using (25) yields

$$\frac{1}{N} \mathbf{q}_{\tau'}^H \mathbf{q}_{t+1} = \frac{\mathbf{q}_{\tau'}^H \mathbf{q}_0}{N} - \frac{\mathbf{q}_{\tau'}^H \eta_t(\mathbf{q}_0 - \mathbf{h}_t)}{N}. \quad (141)$$

The induction hypothesis (42) $\tau < t$ implies that the first term is convergent in the large system limit. In order to prove the existence of (42) for $\tau' \leq \tau = t$, it is sufficient to confirm

$$\frac{1}{N} \mathbf{q}_{\tau'}^H \mathbf{q}_{t+1} \stackrel{\text{a.s.}}{\rightarrow} \zeta_{0,\tau'} - \frac{1}{N} \mathbb{E}_{\mathcal{Z}_t} [\mathbf{q}_{\tau'}^H \eta_t(\mathbf{q}_0 - \mathbf{h}_t^G)] \quad (142)$$

in the large system limit.

Repeating the proof of (99), we use Theorem 1 and (50) for $\tau = t$ to have

$$\frac{\mathbf{q}_{\tau'}^H \mathbf{q}_{t+1}}{N} \stackrel{\text{a.s.}}{\rightarrow} \zeta_{0,\tau'} - \frac{\mathbb{E}_{\mathcal{Z}_t} [\mathbf{q}_{\tau'}^H \eta_t(\mathbf{q}_0 - \tilde{\mathbf{h}}_t)]}{N} \quad (143)$$

in the large system limit, where $\tilde{\mathbf{h}}_t$ is given by (46). Using Theorem 1 and the induction hypothesis (50) repeatedly in the order $\tau = t-1, \dots, 0$, we arrive at (142). Thus, (42) exists for $\tau' \leq \tau = t$. ■

Eq. (56) for $\tau = t$: We shall prove (56) for $\tau = t$. From (132) and Lemma 4, we find

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{h}_t^H \mathbf{q}_{\tau''} \stackrel{\text{a.s.}}{=} \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{m}_t^H \mathbf{b}_{\tau''}, \quad (144)$$

conditioned on Θ and $\mathcal{X}_{t,t+1}$ for $\tau'' \leq t$, which is almost surely equal to zero, because of (53) for $\tau = t$. Thus, (56) holds for $\tau'' \leq \tau = t$.

For $\tau'' = t+1$, we use (25) to have

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{h}_t^H \mathbf{q}_{t+1} \stackrel{\text{a.s.}}{=} - \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{h}_t^H \eta_t(\mathbf{q}_0 - \mathbf{h}_t), \quad (145)$$

where we have used (56) for $\tau'' = 0$ and $\tau = t$. It is possible to prove

$$\begin{aligned} & \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{h}_t^H \mathbf{q}_{t+1} \\ & \stackrel{\text{a.s.}}{=} - \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbb{E}_{\mathcal{Z}_t} [(\mathbf{h}_t^G)^H \eta_t(\mathbf{q}_0 - \mathbf{h}_t^G)], \end{aligned} \quad (146)$$

by repeating the proof of (142).

Let us prove that the RHS of (146) is equal to zero. Since \mathbf{h}_t^G has i.i.d. proper complex Gaussian elements with vanishing mean in the large system limit and variance $v_{A \rightarrow B}^t$, we use Lemma 2 to find that the RHS of (146) is equal to zero. Thus, (56) holds for $\tau = t$. ■

Eqs. (58) and (59) for $\tau = t$: We repeat the proof for the case $\tau = 0$ to obtain (58) and (59) for $\tau = t$. ■

Eqs. (43) and (44) for $\tau = t$: We only prove (44) for $\tau = t$, since (43) can be proved in the same manner. The induction hypothesis (44) for $\tau < t$ implies that (44) holds for $\tau = t$ if $\liminf_{M=\delta N \rightarrow \infty} N^{-1} \|\mathbf{q}_{t+1}^\perp\|^2$ converges almost surely to a strictly positive constant. By repeating the proof of (112) for $\tau = 0$, we obtain

$$\liminf_{M=\delta N \rightarrow \infty} \frac{1}{N} \|\mathbf{q}_{t+1}^\perp\|^2 \stackrel{\text{a.s.}}{\geq} \mathbb{E}[\mathbb{V}_{z_{t,0}}[\eta_t(\tilde{x}_{0,A \rightarrow B}^t)]], \quad (147)$$

which is strictly positive. Thus, (44) holds for $\tau = t$. ■

APPENDIX A PROOF OF LEMMA 1

We use the properties of moments for Haar matrices.

Lemma 6: [31, Lemma 4.2.2, Proposition 4.2.3] Suppose that $\mathbf{V} \in \mathcal{U}_N$ is a Haar matrix. Then, all moments of the elements $\{V_{nm}\}$ in \mathbf{V} up to the 4th order are zero, with the exception of

$$\mathbb{E}[|V_{nm}|^2] = \frac{1}{N}, \quad (148)$$

$$\mathbb{E}[|V_{nm}|^2 |V_{n'm'}|^2] = \begin{cases} \frac{2}{N(N+1)} & \text{for } n = n' \text{ and } m = m', \\ \frac{1}{N^2-1} & \text{for } n \neq n' \text{ and } m \neq m', \\ \frac{1}{N(N+1)} & \text{otherwise,} \end{cases} \quad (149)$$

$$\begin{aligned} & \mathbb{E}[V_{nm} V_{n'm'} V_{nm'}^* V_{n'm}^*] \\ & = -\frac{1}{N(N^2-1)} \text{ for } n \neq n' \text{ and } m \neq m'. \end{aligned} \quad (150)$$

We shall prove Lemma 1. We only present the proof of the latter statement, since the former statement can be proved more straightforwardly. Without loss of generality, we can assume that \mathbf{D} is diagonal, since \mathbf{V} is bi-unitarily invariant. We represent $S_N = \mathbf{b}^H \mathbf{V}^H \mathbf{D} \mathbf{V} \mathbf{a}$ as

$$S_N = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \sum_{m'=0}^{N-1} V_{nm} V_{nm'}^* D_{nn} b_{m'}^* a_m. \quad (151)$$

Using Lemma 6 to evaluate the expectation of S_N over \mathbf{V} yields

$$\frac{1}{N}\mathbb{E}_{\mathbf{V}}[S_N] = \frac{1}{N}\text{Tr}(\mathbf{D})\frac{1}{N}\mathbf{b}^H\mathbf{a} \xrightarrow{\text{a.s.}} C \lim_{N \rightarrow \infty} \frac{1}{N}\text{Tr}(\mathbf{D}), \quad (152)$$

as $N \rightarrow \infty$, because the existence of $\lim_{N \rightarrow \infty} N^{-1}\text{Tr}(\mathbf{D}^2)$ and Hölder's inequality imply that $N^{-1}\text{Tr}(\mathbf{D})$ is convergent as $N \rightarrow \infty$. From Theorem 1 it is sufficient to prove that the variance of S_N is $\mathcal{O}(N^\gamma)$ for some $\gamma < 2$.

We evaluate the second moment of S_N as

$$\mathbb{E}_{\mathbf{V}}[|S_N|^2] = \sum_{n, \tilde{n}, m, \tilde{m}, m', \tilde{m}'} \mathbb{E}[V_{nm}V_{\tilde{n}\tilde{m}'}V_{nm'}^*V_{\tilde{n}\tilde{m}}^*] \cdot D_{nn}D_{\tilde{n}\tilde{n}}a_m b_{m'}^* a_{\tilde{m}}^* b_{\tilde{m}}. \quad (153)$$

Using Lemma 6 yields

$$\begin{aligned} \mathbb{E}_{\mathbf{V}}[|S_N|^2] = & -\frac{1}{N(N^2-1)} \sum_{n \neq \tilde{n}} \sum_{m \neq m'} D_{nn}D_{\tilde{n}\tilde{n}}|a_m|^2|b_{m'}|^2 \\ & + \frac{1}{N^2-1} \sum_{n \neq \tilde{n}} \sum_{m \neq \tilde{m}} D_{nn}D_{\tilde{n}\tilde{n}}a_m b_{m'}^* a_{\tilde{m}}^* b_{\tilde{m}} \\ & + \frac{1}{N(N+1)} \sum_{n \neq \tilde{n}} \sum_m D_{nn}D_{\tilde{n}\tilde{n}}|a_m|^2|b_m|^2 \\ & + \frac{1}{N(N+1)} \sum_n \sum_{m \neq \tilde{m}} D_{nn}^2 a_m b_{m'}^* a_{\tilde{m}}^* b_{\tilde{m}} \\ & + \frac{1}{N(N+1)} \sum_n \sum_{m \neq m'} D_{nn}^2 |a_m|^2|b_{m'}|^2 \\ & + \frac{2}{N(N+1)} \sum_{n, m} D_{nn}^2 |a_m|^2|b_m|^2. \end{aligned} \quad (154)$$

Thus, the variance of S_N over \mathbf{V} is given by

$$\begin{aligned} \mathbb{V}_{\mathbf{V}}[S_N] = & \frac{\|\mathbf{a}\|^2\|\mathbf{b}\|^2}{N^2-1} \left(\text{Tr}(\mathbf{D}^2) - \frac{\text{Tr}^2(\mathbf{D})}{N} \right) \\ & - \frac{|\mathbf{b}^H\mathbf{a}|^2}{N(N^2-1)} \text{Tr}(\mathbf{D}^2) + \frac{|\mathbb{E}_{\mathbf{V}}[S_N]|^2}{N^2-1}, \end{aligned} \quad (155)$$

where we have used $\mathbb{E}_{\mathbf{V}}[S_N] = N^{-1}\text{Tr}(\mathbf{D})\mathbf{b}^H\mathbf{a}$. Note that the term $\sum_{n=0}^{N-1}|a_n|^2|b_n|^2$ vanishes for any N . As a result, we need no assumptions on $\sum_{n=0}^{N-1}|a_n|^4$ for the case $\mathbf{a} = \mathbf{b}$.

We use the continuous mapping theorem and the assumptions on \mathbf{a} , \mathbf{b} , and \mathbf{D} to arrive at

$$\frac{\mathbb{V}_{\mathbf{V}}[S_N]}{N} \xrightarrow{\text{a.s.}} \frac{1}{N}\text{Tr}(\mathbf{D}^2) - \left(\frac{1}{N}\text{Tr}(\mathbf{D}) \right)^2 < \infty \quad (156)$$

as $N \rightarrow \infty$, which implies Lemma 1.

APPENDIX B DERIVATION OF MESSAGE-PASSING

EP [20], [29] provides a framework for deriving MP algorithms that calculate the marginal posterior distribution $p(x_n|\mathbf{y}, \mathbf{A}) = \int p(\mathbf{x}|\mathbf{y}, \mathbf{A})d\mathbf{x}_{\setminus n}$, in which $\mathbf{x}_{\setminus n}$ is the vector obtained by eliminating x_n from \mathbf{x} . We consider the large system limit to derive an MP algorithm, which coincides with the algorithm derived in a heuristic manner [27].

We approximate the marginal posterior distribution $p(x_n|\mathbf{y}, \mathbf{A})$ by a tractable probability density function (pdf) $q_A(x_n) = \int q_A(\mathbf{x})d\mathbf{x}_{\setminus n}$, given by

$$q_A(\mathbf{x}) \propto p(\mathbf{y}|\mathbf{A}, \mathbf{x}) \prod_{n=0}^{N-1} q_{B \rightarrow A}(x_n). \quad (157)$$

In (157), the notation $f(\mathbf{x}) \propto g(\mathbf{x})$ means that there is a positive constant C such that $f(\mathbf{x}) = Cg(\mathbf{x})$ holds. Furthermore, $q_{B \rightarrow A}(x_n)$ is a conjugate prior to the likelihood $p(\mathbf{y}|\mathbf{A}, \mathbf{x})$. When the noise vector \mathbf{w} in (1) is regarded as a CSCG random vector with covariance $\sigma^2\mathbf{I}_M$, the conjugate prior $q_{B \rightarrow A}(x_n)$ is proper complex Gaussian,

$$q_{B \rightarrow A}(x_n) \propto \exp\left(-\frac{|x_n - x_{n, B \rightarrow A}|^2}{v_{B \rightarrow A}}\right), \quad (158)$$

where $x_{n, B \rightarrow A}$ and $v_{B \rightarrow A}$ are the mean and variance of $q_{B \rightarrow A}(x_n)$, respectively. In order to derive the MP algorithm proposed in [27], we have selected the identical variance $v_{B \rightarrow A}$ for all n , while Céspedes *et al.* [20] selected different values for different n to improve the performance for finite-sized systems.

We first evaluate the marginal pdf $q_A(x_n)$ in the large system limit, defined via (157). Since the conjugate prior (158) has been selected, the joint pdf $q_A(\mathbf{x})$ is also Gaussian.

$$q_A(\mathbf{x}) \propto \exp\{(\mathbf{x} - \mathbf{x}_A)^H \mathbf{V}_A^{-1}(\mathbf{x} - \mathbf{x}_A)\}, \quad (159)$$

where the mean and covariance are given by

$$\mathbf{x}_A = \mathbf{x}_{B \rightarrow A} + \frac{1}{\sigma^2} \mathbf{V}_A \mathbf{A}^H (\mathbf{y} - \mathbf{A} \mathbf{x}_{B \rightarrow A}), \quad (160)$$

$$\mathbf{V}_A = \left(\frac{1}{v_{B \rightarrow A}} \mathbf{I}_N + \frac{1}{\sigma^2} \mathbf{A}^H \mathbf{A} \right)^{-1}, \quad (161)$$

respectively. Using the matrix inversion lemma, it is possible to show that (160) and (161) reduce to

$$\mathbf{x}_A = \mathbf{x}_{B \rightarrow A} + v_{B \rightarrow A} \mathbf{A}^H \mathbf{\Xi}^{-1} (\mathbf{y} - \mathbf{A} \mathbf{x}_{B \rightarrow A}), \quad (162)$$

$$[\mathbf{V}_A]_{n,n} = v_{B \rightarrow A} - \mathbf{a}_n^H \mathbf{\Xi}^{-1} \mathbf{a}_n v_{B \rightarrow A}^2, \quad (163)$$

respectively, with

$$\mathbf{\Xi} = \sigma^2 \mathbf{I}_M + v_{B \rightarrow A} \mathbf{A} \mathbf{A}^H. \quad (164)$$

We shall prove that $\mathbf{a}_n^H \mathbf{\Xi}^{-1} \mathbf{a}_n$ converges almost surely to $\gamma(v_{B \rightarrow A})^{-1}$ in the large system limit for all n , in which $\gamma(v_{B \rightarrow A})$ is given by (13). Applying the SVD (7) to $\mathbf{a}_n^H \mathbf{\Xi}^{-1} \mathbf{a}_n$, defined via (164), we have

$$\mathbf{a}_n^H \mathbf{\Xi}^{-1} \mathbf{a}_n = \mathbf{e}_n^H \mathbf{V} \mathbf{D} \mathbf{V}^H \mathbf{e}_n. \quad (165)$$

with

$$\mathbf{D} = \begin{pmatrix} \Sigma \\ \mathbf{O} \end{pmatrix} (\sigma^2 \mathbf{I}_M + v_{B \rightarrow A} \Sigma^2)^{-1} (\Sigma, \mathbf{O}). \quad (166)$$

In (165), \mathbf{e}_n denotes the n th column of \mathbf{I}_N . Thus, Lemma 1 and Assumption 2 imply that $\mathbf{a}_n^H \mathbf{\Xi}^{-1} \mathbf{a}_n$ converges almost surely to $\gamma(v_{B \rightarrow A})^{-1}$ in the large system limit.

This observation indicates that for any n the diagonal element (163) converges almost surely to

$$v_A = v_{B \rightarrow A} - \gamma^{-1}(v_{B \rightarrow A})v_{B \rightarrow A}^2 \quad (167)$$

in the large system limit. Thus, the marginal pdf $q_A(x_n) = \int q_A(\mathbf{x}) d\mathbf{x}_{\setminus n}$ is the proper complex Gaussian pdf with mean $x_{n,A} = [\mathbf{x}_A]_n$ and variance v_A , i.e.

$$q_A(x_n) \propto \exp\left(-\frac{|x_n - x_{n,A}|^2}{v_A}\right). \quad (168)$$

In order to present a crucial step in EP, we define the extrinsic pdf of x_n as

$$q_{A \rightarrow B}(x_n) \propto \frac{q_A(x_n)}{q_{B \rightarrow A}(x_n)}. \quad (169)$$

Let $x_{n,B}$ and $v_{n,B}$ denote the mean and variance of x_n with respect to the pdf $p_B(x_n) \propto q_{A \rightarrow B}(x_n)p(x_n)$. The crucial step in EP is to update the message $q_{B \rightarrow A}(x_n)$ so as to satisfy the moment matching conditions [29],

$$\mathbb{E}_{q_B}[x_n] = x_{n,B}, \quad (170)$$

$$\mathbb{V}_{q_B}[x_n] = \frac{1}{N} \sum_{n=0}^{N-1} v_{n,B} \equiv v_B, \quad (171)$$

where the expectations are taken with respect to

$$q_B(x_n) \propto q_{A \rightarrow B}(x_n)q_{B \rightarrow A}^{\text{new}}(x_n). \quad (172)$$

In (172), the updated pdf $q_{B \rightarrow A}^{\text{new}}(x_n)$ is given by

$$q_{B \rightarrow A}^{\text{new}}(x_n) \propto \exp\left(-\frac{|x_n - x_{n,B \rightarrow A}^{\text{new}}|^2}{v_{B \rightarrow A}^{\text{new}}}\right). \quad (173)$$

We first derive module A. Using (158) and (168), we find that the extrinsic pdf (169) reduces to

$$q_{A \rightarrow B}(x_n) \propto \exp\left(-\frac{|x_n - x_{n,A \rightarrow B}|^2}{v_{A \rightarrow B}}\right), \quad (174)$$

with

$$x_{n,A \rightarrow B} = v_{A \rightarrow B} \left(\frac{x_{n,A}}{v_A} - \frac{x_{n,B \rightarrow A}}{v_{B \rightarrow A}} \right), \quad (175)$$

$$\frac{1}{v_{A \rightarrow B}} = \frac{1}{v_A} - \frac{1}{v_{B \rightarrow A}}. \quad (176)$$

Substituting (167) into (176) yields

$$v_{A \rightarrow B} = \gamma(v_{B \rightarrow A}) - v_{B \rightarrow A}, \quad (177)$$

which results in the update rule (10). Similarly, Applying (162), (167), (176), and (177) to (175), we arrive at

$$\mathbf{x}_{A \rightarrow B} = \mathbf{x}_{B \rightarrow A} + \gamma(v_{B \rightarrow A}) \mathbf{A}^H \mathbf{\Xi}^{-1} (\mathbf{y} - \mathbf{A} \mathbf{x}_{B \rightarrow A}), \quad (178)$$

which implies the update rule (9).

We next evaluate the moment matching conditions (170) and (171) to derive module B. Substituting (173) and (174) into (172) yields

$$q_B(x_n) \propto \exp\left(-\frac{|x_n - \tilde{x}_{n,B}|^2}{\tilde{v}_B}\right), \quad (179)$$

with

$$\tilde{x}_{n,B} = \tilde{v}_B \left(\frac{x_{n,A \rightarrow B}}{v_{A \rightarrow B}} + \frac{x_{n,B \rightarrow A}^{\text{new}}}{v_{B \rightarrow A}^{\text{new}}} \right), \quad (180)$$

$$\frac{1}{\tilde{v}_B} = \frac{1}{v_{A \rightarrow B}} + \frac{1}{v_{B \rightarrow A}^{\text{new}}}. \quad (181)$$

Using the moment matching conditions (170) and (171), we arrive at the update rules (17) and (18) in module B,

$$\mathbf{x}_{B \rightarrow A}^{\text{new}} = v_{B \rightarrow A}^{\text{new}} \left(\frac{\mathbf{x}_B}{v_B} - \frac{\mathbf{x}_{A \rightarrow B}}{v_{A \rightarrow B}} \right), \quad (182)$$

$$\frac{1}{v_{B \rightarrow A}^{\text{new}}} = \frac{1}{v_B} - \frac{1}{v_{A \rightarrow B}}. \quad (183)$$

Note that v_B given in (171) converges almost surely to its expectation in the large system limit.

APPENDIX C PROOF OF LEMMA 2

We utilize the following technical lemma:

Lemma 7: We define the cumulant generating function $\chi_t : \mathbb{C} \rightarrow \mathbb{R}$ of the posterior distribution of x_n as

$$\chi_t(z) = \frac{v_{A \rightarrow B}^t}{2} \ln \mathbb{E}_{x_n} \left[\exp \left(-\frac{|z - x_n|^2}{v_{A \rightarrow B}^t} \right) \right] + \frac{|z|^2}{2}. \quad (184)$$

Then, χ_t is twice continuously differentiable with respect to $\Re[z]$ and $\Im[z]$, and satisfies

$$\frac{\partial \chi_t}{\partial \Re[z]} = \Re[\tilde{\eta}_t(z)], \quad \frac{\partial \chi_t}{\partial \Im[z]} = \Im[\tilde{\eta}_t(z)], \quad (185)$$

$$\frac{\partial^2 \chi_t}{\partial \Re[z]^2} + \frac{\partial^2 \chi_t}{\partial \Im[z]^2} = \frac{2}{v_{A \rightarrow B}^t} \mathbb{E} [|x_n - \tilde{\eta}_t(z)|^2 | z]. \quad (186)$$

Proof: The former statement follows from Assumption 1 and the dominated convergence theorem. The latter statement is obtained by calculating the derivatives of χ_t directly. ■

Let us prove Lemma 2. We use (19) to obtain

$$\begin{aligned} & \mathbb{E}_{z_n^t} [(z_n^t)^* \eta_t(x_n + \epsilon + z_n^t)] \\ &= v_{B \rightarrow A}^{t+1} \left(\frac{\mathbb{E}_z [(z_n^t)^* \tilde{\eta}_t(x_n + \epsilon + z_n^t)]}{\text{MMSE}(v_{A \rightarrow B}^t)} - 1 \right). \end{aligned} \quad (187)$$

Thus, it is sufficient to prove (21).

For notational convenience, we write z_n^t as z . By definition, we have

$$z^* \tilde{\eta}_t = \Re[z] \Re[\tilde{\eta}_t] + \Im[z] \Im[\tilde{\eta}_t] + i(\Re[z] \Im[\tilde{\eta}_t] - \Im[z] \Re[\tilde{\eta}_t]). \quad (188)$$

Since $\Re[z]$ and $\Im[z]$ are independent Gaussian random variables with zero-mean and variance $v_{A \rightarrow B}^t/2$, using Stein's lemma [37] yields

$$\begin{aligned} \mathbb{E}_z [z^* \tilde{\eta}_t] &= \frac{v_{A \rightarrow B}^t}{2} \mathbb{E}_z \left[\frac{\partial \Re[\tilde{\eta}_t]}{\partial \Re[z]} + \frac{\partial \Im[\tilde{\eta}_t]}{\partial \Im[z]} \right] \\ &\quad + \frac{i v_{A \rightarrow B}^t}{2} \mathbb{E}_z \left[\frac{\partial \Im[\tilde{\eta}_t]}{\partial \Re[z]} - \frac{\partial \Re[\tilde{\eta}_t]}{\partial \Im[z]} \right]. \end{aligned} \quad (189)$$

Applying Lemma 7, we obtain

$$\mathbb{E}_z [z^* \tilde{\eta}_t(x_n + \epsilon + z)] = \mathbb{E}_z [|x_n - \tilde{\eta}_t(x_n + \epsilon + z)|^2], \quad (190)$$

which tends to $\text{MMSE}(v_{A \rightarrow B}^t)$ as $\epsilon \rightarrow 0$.

APPENDIX D PROOF OF LEMMA 3

We first prove the following technical lemma:

Lemma 8: Suppose that $\tilde{V} \in \mathcal{U}_{N-t-t'}$ is a Haar matrix and independent of V , and that $V_{10}^{t,t'}$ and $V_{01}^{t,t'}$ in (60) are full rank for $t > 0$ and $t' > 0$. Let $\mathcal{V}_{0,1} = \{V_{01}^{0,1}\}$ and $\mathcal{V}_{t,t'} = \{V_{00}^{t,t'}, V_{01}^{t,t'}, V_{10}^{t,t'}\}$ for $t > 0$ and $t' > 0$. Then, the distribution of $V_{11}^{t,t'}$ conditioned on $\mathcal{V}_{t,t'}$ satisfies

$$V_{11}^{t,t'} | \mathcal{V}_{t,t'} \sim \bar{V}_{11}^{t,t'} + \Phi_{V_{10}^{t,t'}}^\perp \tilde{V} (\Psi_{V_{01}^{t,t'}}^\perp)^H, \quad (191)$$

where $\bar{V}_{11}^{t,t'}$ for $t > 0$ and $t' > 0$ is given in Lemma 3, while the convention $\bar{V}_{11}^{0,1} = O$ is introduced.

Proof: Conditioning on $\mathcal{V}_{t,t'}$ is considered throughout the proof of Lemma 8. We first consider the case $t > 0$ and $t' > 0$. For notational convenience, we omit the scripts t and t' . Define two unitary matrices $\Phi \in \mathcal{U}_N$ and $\Psi \in \mathcal{U}_N$ as

$$\Phi = \begin{pmatrix} \Phi_{V_{01}} & O \\ O & \Phi_{V_{10}} \end{pmatrix}, \quad (192)$$

$$\Psi = \begin{pmatrix} \Psi_{V_{10}} & O \\ O & \Psi_{V_{01}} \end{pmatrix}. \quad (193)$$

Using the SVDs (61) and (63) to calculate $\hat{V} = \Phi^H V \Psi$ yields

$$\hat{V} = \begin{pmatrix} \hat{V}_{00} & (\Sigma_{V_{01}}, O) \\ (\Sigma_{V_{10}}, O) & \hat{V}_{11} \end{pmatrix}, \quad (194)$$

with

$$\hat{V}_{00} = \Phi_{V_{01}}^H V_{00} \Psi_{V_{10}} \in \mathbb{C}^{t' \times t}, \quad (195)$$

$$\hat{V}_{11} = \Phi_{V_{10}}^H V_{11} \Psi_{V_{01}} \in \mathbb{C}^{(N-t') \times (N-t)}. \quad (196)$$

Note that $\hat{V} \in \mathcal{U}_N$ holds.

We next confirm that \hat{V}_{11} has the following block structure:

$$\hat{V}_{11} \sim \begin{pmatrix} \tilde{V}_{11} & O \\ O & \tilde{V} \end{pmatrix}. \quad (197)$$

In (197), $\tilde{V}_{11} \in \mathbb{C}^{t \times t'}$ is given by

$$\tilde{V}_{11} = -\Sigma_{V_{10}} \Psi_{V_{10}}^H V_{00}^H \Phi_{V_{01}} \Sigma_{V_{01}}^{-1} \quad (198)$$

$$= -\Sigma_{V_{10}}^{-1} \Psi_{V_{10}}^H V_{00}^H \Phi_{V_{01}} \Sigma_{V_{01}}. \quad (199)$$

Furthermore, $\tilde{V} \in \mathcal{U}_{N-t-t'}$ is a Haar matrix and independent of V . Note that $\Sigma_{V_{10}} \in \mathbb{C}^{t \times t}$ and $\Sigma_{V_{01}} \in \mathbb{C}^{t' \times t'}$ are invertible, since V_{10} and V_{01} are full rank.

The structure (198) in the upper-left block follows from the orthogonality between the first $t + t'$ rows of \hat{V} . Similarly, (199) is due to the orthogonality between the first $t + t'$ columns of \hat{V} . It is possible to confirm the consistency of the two expressions (198) and (199), by using the properties $V_{00}^H V_{00} + V_{10}^H V_{10} = I_t$ and $V_{00} V_{00}^H + V_{01} V_{01}^H = I_{t'}$.

The structures in the off-diagonal blocks follow from the orthogonality between the first t columns and the last $(N - t - t')$ columns of \hat{V} and between the first t' rows and the last $(N - t - t')$ rows. The structure in the bottom-right block is due to the orthonormality between the last $N - t - t'$ columns.

In order to complete the proof, we use (196) and (197) to obtain

$$V_{11} | \mathcal{V} \sim \Phi_{V_{10}}^\parallel \tilde{V}_{11} (\Psi_{V_{01}}^\parallel)^H + \Phi_{V_{10}}^\perp \tilde{V} (\Psi_{V_{01}}^\perp)^H. \quad (200)$$

Applying (198) and (199), as well as (61) and (63), we find that the first term reduces to (71) and (72), respectively. Thus, (191) holds.

We next consider the case $t = 0$ and $t' = 1$, and omit the scripts t and t' . Let

$$\Phi = \begin{pmatrix} \Phi_{V_{01}} & O \\ O & I_{N-1} \end{pmatrix}. \quad (201)$$

Using the SVD (63) to calculate $\hat{V} = \Phi^H V \Psi_{V_{01}}$ yields

$$\hat{V} = \begin{pmatrix} (\Sigma_{V_{01}}, O) \\ V_{11} \Psi_{V_{01}} \end{pmatrix}. \quad (202)$$

By using the orthogonality between the rows of \hat{V} , it is straightforward to confirm that $\hat{V} \in \mathcal{U}_N$ has the following structure:

$$\hat{V} = \begin{pmatrix} \Sigma_{V_{01}} & O \\ O & \tilde{V} \end{pmatrix}, \quad (203)$$

where $\tilde{V} \in \mathcal{U}_{N-1}$ is a Haar matrix and independent of V . Thus, we use $V = \Phi \hat{V} \Psi_{V_{01}}^H$ and the convention $\Phi_{V_{01}}^\perp = I_{N-1}$ to arrive at (191). ■

Let us prove Lemma 3. We first assume $t > 0$ and $t' > 0$. Substituting the SVDs (62) and (64) into (28) and (30) yields

$$\hat{V}^H \begin{pmatrix} \Sigma_{Q_{t'}} \\ O \end{pmatrix} = \Phi_{M_t}^H B_{t'} \Psi_{Q_{t'}}, \quad (204)$$

$$\hat{V} \begin{pmatrix} \Sigma_{M_t} \\ O \end{pmatrix} = \Phi_{Q_{t'}}^H H_t \Psi_{M_t}, \quad (205)$$

with

$$\hat{V} = \Phi_{Q_{t'}}^H V \Phi_{M_t}. \quad (206)$$

Since $\Sigma_{Q_{t'}}$ and Σ_{M_t} are assumed to be invertible, conditioning on Θ and $\mathcal{X}_{t,t'}$ is equivalent to observing the first t columns and the first t' rows of \hat{V} . Note that $\hat{V} \sim V$ holds, since the Haar matrix V is bi-unitarily invariant. Thus, we have

$$\hat{V} | \Theta, \mathcal{X}_{t,t'} \sim \begin{pmatrix} \hat{V}_{00} & \hat{V}_{01} \\ \hat{V}_{10} & \hat{V}_{11} | \hat{\mathcal{V}} \end{pmatrix}, \quad (207)$$

with $\hat{\mathcal{V}} = \{\hat{V}_{00}, \hat{V}_{01}, \hat{V}_{10}\}$, partitioned in the same manner as in (60).

We shall prove that (207) is equivalent to (69). From (204) and (205), we have

$$\hat{V}_{00} = \Sigma_{Q_{t'}}^{-1} \Psi_{Q_{t'}}^H B_{t'}^H \Phi_{M_t}^\parallel \quad (208)$$

$$= (\Phi_{Q_{t'}}^\parallel)^H H_t \Psi_{M_t} \Sigma_{M_t}^{-1}, \quad (209)$$

$$\hat{V}_{01} = \Sigma_{Q_{t'}}^{-1} \Psi_{Q_{t'}}^H B_{t'}^H \Phi_{M_t}^\perp, \quad (210)$$

$$\hat{V}_{10} = (\Phi_{Q_{t'}}^\perp)^H H_t \Psi_{M_t} \Sigma_{M_t}^{-1}. \quad (211)$$

It is possible to confirm the consistency of (208) and (209), by using the condition $B_{t'}^H M_t = Q_{t'}^H H_t$. Using the SVDs (62) and (64), we find that the RHSs of (208)–(211) reduce to the RHSs of (65)–(68). Since $Q_{t'}$ and M_t are full rank,

(28) and (30) imply that $\mathbf{B}_{t'}$ and \mathbf{H}_t are so. Thus, (210) and (211) are full rank. Using Lemma 8 and (206), we find that (207) reduces to (69).

We next consider the case $t = 0$ and $t' = 1$. In this case, the SVD (64) reduces to

$$\mathbf{Q}_1 = \Phi_{\mathbf{Q}_1} \begin{pmatrix} \|q_0\| \\ \mathbf{0} \end{pmatrix}, \quad (212)$$

where the first column of $\Phi_{\mathbf{Q}_1}$ is equal to $q_0/\|q_0\|$. Since $\mathbf{V}^H \mathbf{Q}_1 = \mathbf{b}_0$ is the only linear constraint, we have

$$\hat{\mathbf{V}}^H \begin{pmatrix} \|q_0\| \\ \mathbf{0} \end{pmatrix} = \mathbf{b}_0, \quad (213)$$

instead of (204), with $\hat{\mathbf{V}} = \Phi_{\mathbf{Q}_1}^H \mathbf{V}$. Expression (213) implies that conditioning on Θ and $\mathcal{X}_{0,1}$ is equivalent to observing the first row of $\hat{\mathbf{V}}$. Thus, we have

$$\hat{\mathbf{V}}|_{\Theta, \mathcal{X}_{0,1}} \sim \begin{pmatrix} \hat{\mathbf{V}}_{01} \\ \hat{\mathbf{V}}_{11} |_{\hat{\mathbf{V}}_{01}} \end{pmatrix}, \quad (214)$$

with $\hat{\mathbf{V}}_{01} = \mathbf{b}_0^H/\|q_0\|$. Using $\mathbf{V} = \Phi_{\mathbf{Q}_1} \hat{\mathbf{V}}$, Lemma 8, and the convention $\Phi_{\mathbf{M}_0}^\perp = \mathbf{I}_N$, we arrive at (69) for $t = 0$ and $t' = 1$.

APPENDIX E PROOF OF LEMMA 4

For notational convenience, the superscripts t and t' are omitted. We first prove (77), (78) and (80). Using (71), we find that (70) reduces to

$$\begin{aligned} \bar{\mathbf{V}}_{t,t'} &= \Phi_{\mathbf{Q}_{t'}}^\perp (\mathbf{V}_{00}, \mathbf{V}_{01}) \Phi_{\mathbf{M}_t}^H + \Phi_{\mathbf{Q}_{t'}}^\perp \mathbf{V}_{10} (\Phi_{\mathbf{M}_t}^\perp)^H \\ &\quad - \Phi_{\mathbf{Q}_{t'}}^\perp \mathbf{V}_{10} (\mathbf{V}_{01}^\dagger \mathbf{V}_{00})^H (\Phi_{\mathbf{M}_t}^\perp)^H. \end{aligned} \quad (215)$$

Substituting (65) and (67) into the first term on the RHS of (215), we have

$$\Phi_{\mathbf{Q}_{t'}}^\perp (\mathbf{V}_{00}, \mathbf{V}_{01}) \Phi_{\mathbf{M}_t}^H = (\mathbf{B}_{t'} \mathbf{Q}_{t'}^\dagger)^H, \quad (216)$$

where we have used $\mathbf{Q}_{t'}^\dagger \mathbf{P}_{\mathbf{Q}_{t'}}^\perp = \mathbf{Q}_{t'}^\dagger$. Similarly, substituting (68) into the second term yields

$$\Phi_{\mathbf{Q}_{t'}}^\perp \mathbf{V}_{10} (\Phi_{\mathbf{M}_t}^\perp)^H = \mathbf{P}_{\mathbf{Q}_{t'}}^\perp \mathbf{H}_t \mathbf{M}_t^\dagger, \quad (217)$$

since $\mathbf{M}_t^\dagger \mathbf{P}_{\mathbf{M}_t}^\perp = \mathbf{M}_t^\dagger$ holds.

In order to evaluate the last term, we use (67) to calculate \mathbf{V}_{01}^\dagger as

$$\mathbf{V}_{01}^\dagger = (\Phi_{\mathbf{M}_t}^\perp)^H \mathbf{B}_{t'} (\mathbf{B}_{t'}^\dagger \mathbf{P}_{\mathbf{M}_t}^\perp \mathbf{B}_{t'})^{-1} \{(\mathbf{Q}_{t'}^\dagger \Phi_{\mathbf{Q}_{t'}}^\perp)^H\}^{-1}. \quad (218)$$

Substituting (65), (68), and (218) into the last term of (215), and using (216) and (217), we have

$$\bar{\mathbf{V}}_{t,t'} = (\mathbf{B}_{t'} \mathbf{Q}_{t'}^\dagger)^H + \mathbf{P}_{\mathbf{Q}_{t'}}^\perp \mathbf{H}_t \mathbf{\Gamma}_{t,t'}, \quad (219)$$

where $\mathbf{\Gamma}_{t,t'}$ is given by (75).

Expression (77) follows from (219), $\mathbf{B}_{t'} \mathbf{Q}_{t'}^\dagger \mathbf{q}_\tau = \mathbf{b}_\tau$, and $\mathbf{P}_{\mathbf{Q}_{t'}}^\perp \mathbf{q}_\tau = \mathbf{0}$ for $\tau < t'$. Using $\beta_t = \mathbf{Q}_t^\dagger \mathbf{q}_t$, $\mathbf{q}_t^\perp = \mathbf{P}_{\mathbf{Q}_t}^\perp \mathbf{q}_t$, and (219) yields (78). We use (67) and (218) to obtain (80).

We next prove (79) and (81). Using (70) and (72) yields

$$\begin{aligned} \bar{\mathbf{V}}_{t,t'} &= \Phi_{\mathbf{Q}_{t'}} \begin{pmatrix} \mathbf{V}_{00} \\ \mathbf{V}_{10} \end{pmatrix} (\Phi_{\mathbf{M}_t}^\perp)^H + \Phi_{\mathbf{Q}_{t'}}^\perp \mathbf{V}_{01} (\Phi_{\mathbf{M}_t}^\perp)^H \\ &\quad - \Phi_{\mathbf{Q}_{t'}}^\perp (\mathbf{V}_{00} \mathbf{V}_{10}^\dagger)^H \mathbf{V}_{01} (\Phi_{\mathbf{M}_t}^\perp)^H. \end{aligned} \quad (220)$$

To evaluate the last term, we use (68) to obtain

$$\mathbf{V}_{10}^\dagger = (\mathbf{M}_t^\dagger \Phi_{\mathbf{M}_t}^\perp)^{-1} (\mathbf{H}_t^H \mathbf{P}_{\mathbf{Q}_{t'}}^\perp \mathbf{H}_t)^{-1} \mathbf{H}_t^H \Phi_{\mathbf{Q}_{t'}}^\perp. \quad (221)$$

Substituting (66), (67), (68), and (221) into (220), we arrive at

$$\bar{\mathbf{V}}_{t,t'} = \mathbf{H}_t \mathbf{M}_t^\dagger + \Delta_{t,t'}^H \mathbf{B}_{t'}^H \mathbf{P}_{\mathbf{M}_t}^\perp, \quad (222)$$

with $\Delta_{t,t'}$ given by (76), in which we have used $\mathbf{M}_t^\dagger \mathbf{P}_{\mathbf{M}_t}^\perp = \mathbf{M}_t^\dagger$ and $\mathbf{Q}_{t'}^\dagger \mathbf{P}_{\mathbf{Q}_{t'}}^\perp = \mathbf{Q}_{t'}^\dagger$.

Expression (79) follows from (222), $\alpha_t = \mathbf{M}_t^\dagger \mathbf{m}_t$, and $\mathbf{m}_t^\perp = \mathbf{P}_{\mathbf{M}_t}^\perp \mathbf{m}_t$. We use (68) and (221) to obtain (81).

APPENDIX F PROOF OF LEMMA 5

We only prove (87), since (84) can be proved in the same manner. We first prove the case $t = 0$, i.e.

$$\Phi_{\mathbf{Q}_1}^\perp \tilde{\mathbf{V}} \mathbf{a} \xrightarrow{d} \mathbf{z} + o(1) \mathbf{q}_0 \quad (223)$$

conditioned on \mathbf{a} , Θ , and $\mathcal{X}_{0,1}$. For any $\mathbf{V}_0 \in \mathbb{C}^{1 \times (N-1)}$, (223) can be represented as

$$\Phi_{\mathbf{Q}_1}^\perp \tilde{\mathbf{V}} \mathbf{a} = \Phi_{\mathbf{Q}_1} \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N-1} \end{pmatrix} \begin{pmatrix} \mathbf{V}_0 \\ \tilde{\mathbf{V}} \end{pmatrix} \mathbf{a} = (\mathbf{I}_N - \mathbf{P}_{\mathbf{Q}_1}^\perp) \tilde{\mathbf{z}}, \quad (224)$$

with

$$\tilde{\mathbf{z}} = \Phi_{\mathbf{Q}_1} \begin{pmatrix} \mathbf{V}_0 \\ \tilde{\mathbf{V}} \end{pmatrix} \mathbf{a}. \quad (225)$$

In particular, we select an i.i.d. Gaussian vector $\mathbf{V}_0 \in \mathbb{C}^{1 \times (N-1)}$ that is independent of all random variables and has independent CSCG elements with variance $(N-1)^{-1}$. Then, Theorem 2 implies that, for any $k \in \mathbb{N}$, $\tilde{\mathbf{z}}_{\mathcal{N}}$ conditioned on \mathbf{a} , Θ , and $\mathcal{X}_{0,1}$ converges in distribution to $\mathbf{z}_{\mathcal{N}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_k)$ for all subsets $\mathcal{N} \in \mathfrak{N}_k$ in the large system limit.

To complete the proof, we let $\epsilon = (\mathbf{Q}_1^H \mathbf{Q}_1)^{-1} \mathbf{Q}_1^H \tilde{\mathbf{z}}$, and prove that ϵ converges almost surely to zero in the large system limit. The assumptions in Lemma 5 imply that there is a constant $C > 0$ such that

$$|\epsilon| = \left| \left(\frac{1}{N} \mathbf{Q}_1^H \mathbf{Q}_1 \right)^{-1} \frac{\mathbf{Q}_1^H \tilde{\mathbf{z}}}{N} \right| \stackrel{\text{a.s.}}{\leq} C \left| \frac{\mathbf{Q}_1^H \tilde{\mathbf{z}}}{N} \right|. \quad (226)$$

From Theorem 1, it is sufficient to confirm that $N^{-\gamma} \mathbb{V}[\mathbf{Q}_1^H \tilde{\mathbf{z}}]$ is bounded for some $\gamma < 2$. We select $\gamma = 1$ to have

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbb{V}[\mathbf{Q}_1^H \tilde{\mathbf{z}}] = \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \text{Tr}(\mathbf{Q}_1^H \mathbf{Q}_1) < \infty. \quad (227)$$

Thus, we find that ϵ converges almost surely to zero in the large system limit.

We next consider the case $t > 0$. Let $\Phi^\perp = \Phi_{\mathbf{Q}_{t+1}}^\perp \Phi_{\mathbf{V}_{10}^{t,t+1}}^\perp$. Since $\Phi^\perp \in \mathcal{U}_{N \times (N-2t-1)}$ holds, we can construct a unitary matrix $\Phi = (\Phi^\perp, \Phi^\perp) \in \mathcal{U}_N$ by adding $\Phi^\perp \in \mathcal{U}_{N \times (2t+1)}$. Lemma 4 implies $\Phi^\perp (\Phi^\perp)^H = \mathbf{P}_{\mathbf{Q}_{t+1}}^\perp - \mathbf{P}_{\mathbf{P}_{\mathbf{Q}_{t+1}}^\perp}^\perp \mathbf{H}_t$, so that

$\Phi^\parallel(\Phi^\parallel)^H = P_{Q_{t+1}}^\parallel + P_{P_{Q_{t+1}}^\perp H_t}^\parallel$ must hold. Repeating the proof of (224) for $\Phi^\perp \tilde{V}a$, we have

$$\Phi_{Q_{t+1}}^\perp \Phi_{V_{10}^{t,t+1}}^\perp \tilde{V}a \stackrel{d}{\rightarrow} z - (P_{Q_{t+1}}^\parallel + P_{P_{Q_{t+1}}^\perp H_t}^\parallel)z \quad (228)$$

conditioned on a , Θ , and $\mathcal{X}_{t,t+1}$ in the large system limit, where $z_{\mathcal{N}} \sim \mathcal{CN}(\mathbf{0}, I_k)$ holds for all subsets $\mathcal{N} \in \mathfrak{N}_k$. Evaluating the second term in the same manner as in the derivation of (223), we arrive at (87).

REFERENCES

- [1] K. Takeuchi, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," accepted for presentation at 2017 *IEEE Int. Symp. Inf. Theory*. [Online]. Available: <http://arxiv.org/abs/1701.05653>.
- [2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [3] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [4] Y. Wu and S. Verdú, "Rényi information dimension: Fundamental limits of almost lossless analog compression," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3721–3748, Aug. 2010.
- [5] D. J. Thouless, P. W. Anderson, and R. G. Palmer, "Solution of 'solvable model of a spin glass'," *Philos. Mag.*, vol. 35, no. 3, pp. 593–601, 1977.
- [6] Y. Kabashima, "A CDMA multiuser detection algorithm on the basis of belief propagation," *J. Phys. A: Math. Gen.*, vol. 36, no. 43, pp. 11 111–11 121, Oct. 2003.
- [7] K. Takeuchi, T. Tanaka, and T. Kawabata, "Performance improvement of iterative multiuser detection for large sparsely-spread CDMA systems by spatial coupling," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1768–1794, Apr. 2015.
- [8] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.
- [9] G. Reeves and H. D. Pfister, "The replica-symmetric prediction for compressed sensing with Gaussian matrices is exact," in *Proc. 2016 IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016, pp. 665–669.
- [10] J. Barbier, M. Dia, N. Macris, and F. Krzakala, "The mutual information in random linear estimation," in *Proc. 54th Annual Allerton Conf. Commun. Control & Computing*, Urbana-Champaign, IL, USA, Sep. 2016, pp. 625–632.
- [11] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009.
- [12] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.
- [13] S. Kudekar, T. Richardson, and R. Urbanke, "Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 803–834, Feb. 2011.
- [14] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová, "Statistical-physics-based reconstruction in compressed sensing," *Phys. Rev. X*, vol. 2, pp. 021 005–1–18, May 2012.
- [15] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7434–7464, Nov. 2013.
- [16] F. Caltagirone, L. Zdeborová, and F. Krzakala, "On convergence of approximate message passing," in *Proc. 2014 IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jul. 2014, pp. 1812–1816.
- [17] S. Rangan, P. Schniter, and A. Fletcher, "On the convergence of approximate message passing with arbitrary matrices," in *Proc. 2014 IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jul. 2014, pp. 236–240.
- [18] M. Oppor and O. Winther, "Expectation consistent approximate inference," *J. Mach. Learn. Res.*, vol. 6, pp. 2177–2204, Dec. 2005.
- [19] —, "Adaptive and self-averaging Thouless-Anderson-Palmer mean-field theory for probabilistic modeling," *Phys. Rev. E*, vol. 64, no. 5, pp. 056 131–1–14, 2001.
- [20] J. Céspedes, P. M. Olmos, M. Sánchez-Fernández, and F. Perez-Cruz, "Expectation propagation detection for high-order high-dimensional MIMO systems," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2840–2849, Aug. 2014.
- [21] Y. Kabashima and M. Vehkaperä, "Signal recovery using expectation consistent approximation for linear observations," in *Proc. 2014 IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jul. 2014, pp. 226–230.
- [22] A. K. Fletcher, M. Sahraee-Ardakan, S. Rangan, and P. Schniter, "Expectation consistent approximate inference: Generalizations and convergence," in *Proc. 2016 IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016, pp. 190–194.
- [23] B. Çakmak, O. Winther, and B. H. Fleury, "S-AMP: Approximate message passing for general matrix ensembles," in *Proc. 2014 IEEE Inf. Theory Workshop*, Hobart, Australia, Nov. 2014, pp. 192–196.
- [24] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," [Online]. Available: <http://arxiv.org/abs/1610.03082>.
- [25] X. Yuan and J. Ma, "Iterative equalization for MIMO systems: Algorithm design and evolution analysis," in *Proc. 2013 IEEE Wirel. Commun. Netw. Conf.*, Shanghai, China, Apr. 2013, pp. 3974–3979.
- [26] J. Ma, X. Yuan, and L. Ping, "Turbo compressed sensing with partial DFT sensing matrix," *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 158–161, Feb. 2015.
- [27] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol. 5, pp. 2020–2033, Jan. 2017.
- [28] T. Liu, C.-K. Wen, S. Jin, and X. You, "Generalized turbo signal recovery for nonlinear measurements and orthogonal sensing matrices," in *Proc. 2016 IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016, pp. 2883–2887.
- [29] T. P. Minka, "Expectation propagation for approximate Bayesian inference," in *Proc. 17th Conf. Uncertainty Artif. Intell.*, Seattle, WA, USA, Aug. 2001, pp. 362–369.
- [30] K. Takeda, S. Uda, and Y. Kabashima, "Analysis of CDMA systems that are characterized by eigenvalue spectrum," *Europhys. Lett.*, vol. 76, no. 6, pp. 1193–1199, 2006.
- [31] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*. Amer. Math. Soc., 2000.
- [32] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. Hanover, MA USA: Now Publishers Inc., 2004.
- [33] E. Meckes, "Linear functions on the classical matrix groups," *Trans. Amer. Math. Soc.*, vol. 360, no. 10, pp. 5355–5366, Oct. 2008.
- [34] S. Chatterjee and E. Meckes, "Multivariate normal approximation using exchangeable pairs," *ALEA Latin. Amer. J. Prob. Math. Stat.*, vol. 4, pp. 257–283, 2008.
- [35] K. Takeuchi, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," [Online]. Available: <http://arxiv.org/abs/1701.05284>.
- [36] R. Lyons, "Strong laws of large numbers for weakly correlated random variables," *Michigan Math. J.*, vol. 35, no. 3, pp. 353–359, 1988.
- [37] C. Stein, "A bound for the error in the normal approximation to the distribution of a sum of dependent random variables," in *6th Berkeley Symp. Math. Statist. Prob.*, vol. 2, 1972, pp. 583–602.